STIC-Biotech/ChemLib 118892

From: Whiteman, Brian
Sent: Tuesday, April 06, 2004 6:34 PM
To: STIC-Biotech/ChemLib
Subject: seq search

09/610,313
7/5/00
Barnett et al.,

search SEQ ID NO: 30, 31, and 32 against us patent and us patent application databases.

Thanks,
Brian Whiteman
Remsen, 2D14
mail box 2C18
Patent Examiner - Art Unit 1635
United States Patent and Trademark Office
(571) 272-0764

Searcher:_____
Phone:_____
Location:_____
Date Picked Up: 4/7/04
Date Completed: 4/15/04
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences: 3
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.: 0.0 4
WWW/Internet:_____
Other (specify):_____

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:53:16 ; Search time 6311.25 Seconds
(without alignments)
16873.640 Million cell updates/sec

Title: US-09-610-313-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	100.0	2457	6	AX455916	AX455916 Sequence
2	2436.2	99.2	2463	6	AX455915	AX455915 Sequence
3	2415.4	98.3	2469	6	AX455914	AX455914 Sequence
4	2040	83.0	2300	6	BD263705	BD263705 Improved
5	2040	83.0	2300	6	AR373388	AR373388 Sequence
6	2019.2	82.2	2306	6	BD263704	BD263704 Improved
7	2019.2	82.2	2306	6	AR373387	AR373387 Sequence
8	2000.8	81.4	9166	6	AX427930	AX427930 Sequence
9	1998.4	81.3	2312	6	BD263706	BD263706 Improved
10	1998.4	81.3	2312	6	AR373389	AR373389 Sequence
11	1992.4	81.1	9788	6	AX427936	AX427936 Sequence
12	1978.4	80.5	9169	6	AX427931	AX427931 Sequence
13	1975	80.4	9194	6	AX427926	AX427926 Sequence
14	1973.4	80.3	9194	6	AX427925	AX427925 Sequence
15	1973.4	80.3	12411	6	AX427927	AX427927 Sequence
16	1970	80.2	9785	6	AX427938	AX427938 Sequence
17	1966.6	80.0	9167	6	AX427933	AX427933 Sequence
18	1966.6	80.0	9170	6	AX427928	AX427928 Sequence
19	1966.6	80.0	9782	6	AX427935	AX427935 Sequence
20	1966.6	80.0	9783	6	AX427934	AX427934 Sequence
21	1966.6	80.0	9792	6	AX427932	AX427932 Sequence
22	1965	80.0	9189	6	AX427921	AX427921 Sequence
23	1962	79.9	4353	12	AF287353	AF287353 Synthetic
24	1945.4	79.2	3015	6	AX455946	AX455946 Sequence
25	1928.8	78.5	3009	6	AX455987	AX455987 Sequence
26	1915.8	78.0	4319	6	BD263639	BD263639 Improved
27	1915.8	78.0	4319	6	AR373322	AR373322 Sequence
28	1905.2	77.5	4352	12	AF287352	AF287352 Synthetic
29	1878.6	76.5	1965	6	AX455952	AX455952 Sequence
30	1872.8	76.2	2299	6	BD263703	BD263703 Improved
31	1872.8	76.2	2299	6	AR373386	AR373386 Sequence
32	1859.8	75.7	2577	6	AX457088	AX457088 Sequence
33	1852.2	75.4	4343	6	AX306428	AX306428 Sequence
34	1852	75.4	2305	6	BD263702	BD263702 Improved
35	1852	75.4	2305	6	AR373385	AR373385 Sequence
36	1848.2	75.2	4341	6	AX306429	AX306429 Sequence
37	1838.6	74.8	1977	6	AX455954	AX455954 Sequence
38	1825.6	74.3	7897	6	AX427923	AX427923 Sequence
39	1772	72.1	3024	12	AF287355	AF287355 Synthetic
40	1757.2	71.5	9407	6	AX427937	AX427937 Sequence
41	1626.6	66.2	4307	6	AR302563	AR302563 Sequence
42	1626.6	66.2	4307	6	AX019132	AX019132 Sequence
43	1626.6	66.2	4307	6	BD130468	BD130468 Antiviral
44	1607.4	65.4	4307	6	BD268901	BD268901 Anti-vira
45	1607.4	65.4	4307	6	AX035453	AX035453 Sequence

ALIGNMENTS

RESULT 1
AX455916
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)

FEATURES Location/Qualifiers
source 1. .2457
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YMWM"

ORIGIN

Query Match 100.0%; Score 2457; DB 6; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2.4e-248;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
```

```
Db       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457


RESULT 2
AX455915
LOCUS       AX455915               2463 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 31 from Patent WO0204493.
ACCESSION   AX455915
VERSION     AX455915.1  GI:21714900
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 31 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1. .2463
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YM"
ORIGIN

  Query Match           99.2%;  Score 2436.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 3.6e-246;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Db         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Db         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Db         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Db         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Db         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Db         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Db         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Db         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Db         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Db         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy         961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Db         961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Db        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Db        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Db        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy        1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Db        1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
```

```
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
```

```
Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
Qy       2455 TTC 2457
Db       2461 TTC 2463
```

RESULT 3
AX455914
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 30 from Patent WO0204493.
ACCESSION AX455914
VERSION AX455914.1 GI:21714899
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)

FEATURES Location/Qualifiers
source 1. .2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975(+)"

ORIGIN

Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 5.3e-244;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
```

```
Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db      1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy      1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db      1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy      1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
Db      1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy      1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
Db      1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy      1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
Db      1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy      1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
Db      1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy      1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
Db      1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy      1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
Db      1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy      1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Db      1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy      1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
Db      1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy      1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
Db      1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy      1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
Db      1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy      1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
Db      1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy      1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
Db      1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy      2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
Db      2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy      2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
Db      2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy      2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
Db      2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy      2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
Db      2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy      2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
Db      2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy      2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
Db      2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy      2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
Db      2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy      2449 GGTGAATTC 2457
Db      2461 GGTGAATTC 2469
```

RESULT 4
BD263705
LOCUS BD263705 2300 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION BD263705
VERSION BD263705.1 GI:33073473
KEYWORDS JP 2002533124-A/72.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 2300)

```
AUTHORS     Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
TITLE       Improved expression of HIV polypeptides and production of
            virus-like particles
JOURNAL     Patent: JP 2002533124-A 72 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/72
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US     60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YMWM FH
            Key          Location/Qualifiers
            FT   source        1. .2300
            FT                 /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             83.0%;  Score 2040;  DB 6;  Length 2300;
  Best Local Similarity   93.6%;  Pred. No. 1e-204;
  Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500

Qy       1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db       1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy       1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db       1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy       1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db       1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy       1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
```

```
Db       1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy       1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db       1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy       1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db       1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db       1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db       1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db       1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db       2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db       2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2281 GGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 5
AR373388
LOCUS       AR373388                2300 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 83 from patent US 6602705.
ACCESSION   AR373388
VERSION     AR373388.1  GI:40075491
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 83 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2300
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             83.0%;  Score 2040;  DB 6;  Length 2300;
  Best Local Similarity   93.6%;  Pred. No. 1e-204;
  Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;
```

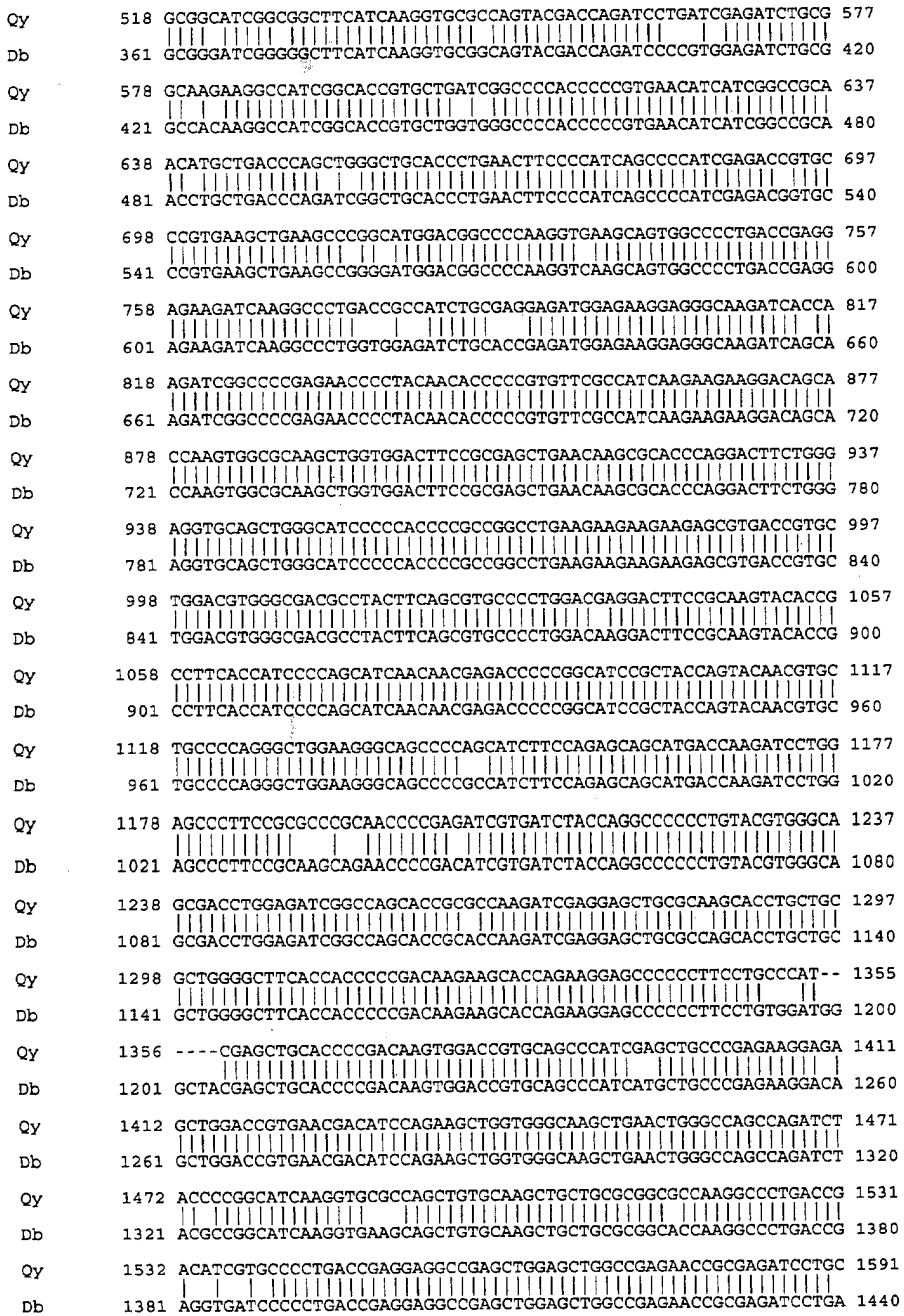
```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
```

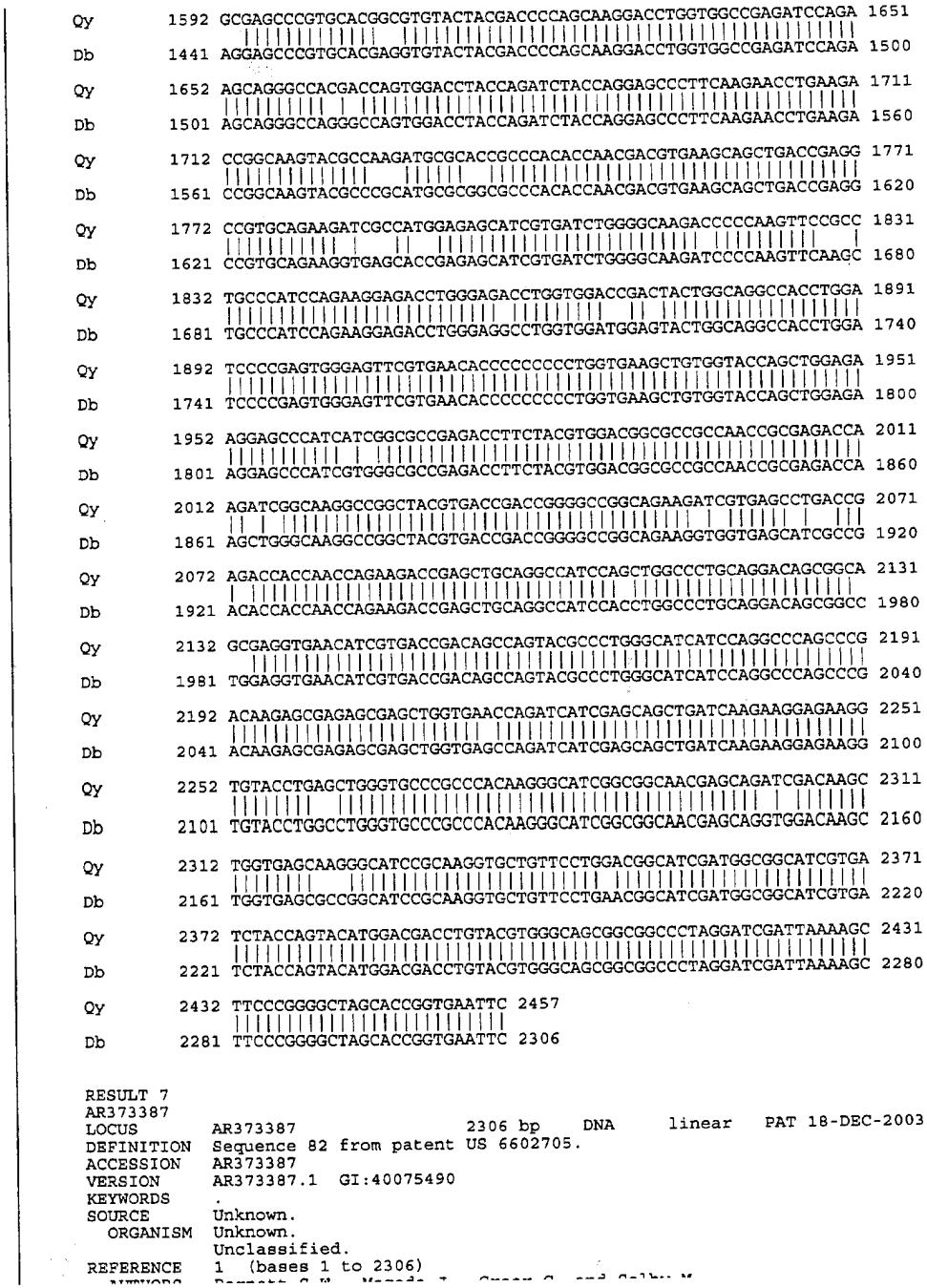
```
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500

Qy       1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db       1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy       1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db       1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy       1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db       1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy       1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db       1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy       1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db       1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy       1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db       1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db       1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db       1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db       1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db       2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db       2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2281 GGGCTAGCACCGGTGAATTC 2300
```

RESULT 6
BD263704

```
LOCUS       BD263704                2306 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263704
VERSION     BD263704.1  GI:33073472
KEYWORDS    JP 2002533124-A/71.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 71 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/71
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH   Key
                     Location/Qualifiers
            FT   source          1. .2306
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN
```

```
  Query Match           82.2%;  Score 2019.2;  DB 6;  Length 2306;
  Best Local Similarity   93.2%;  Pred. No. 1.5e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360
```

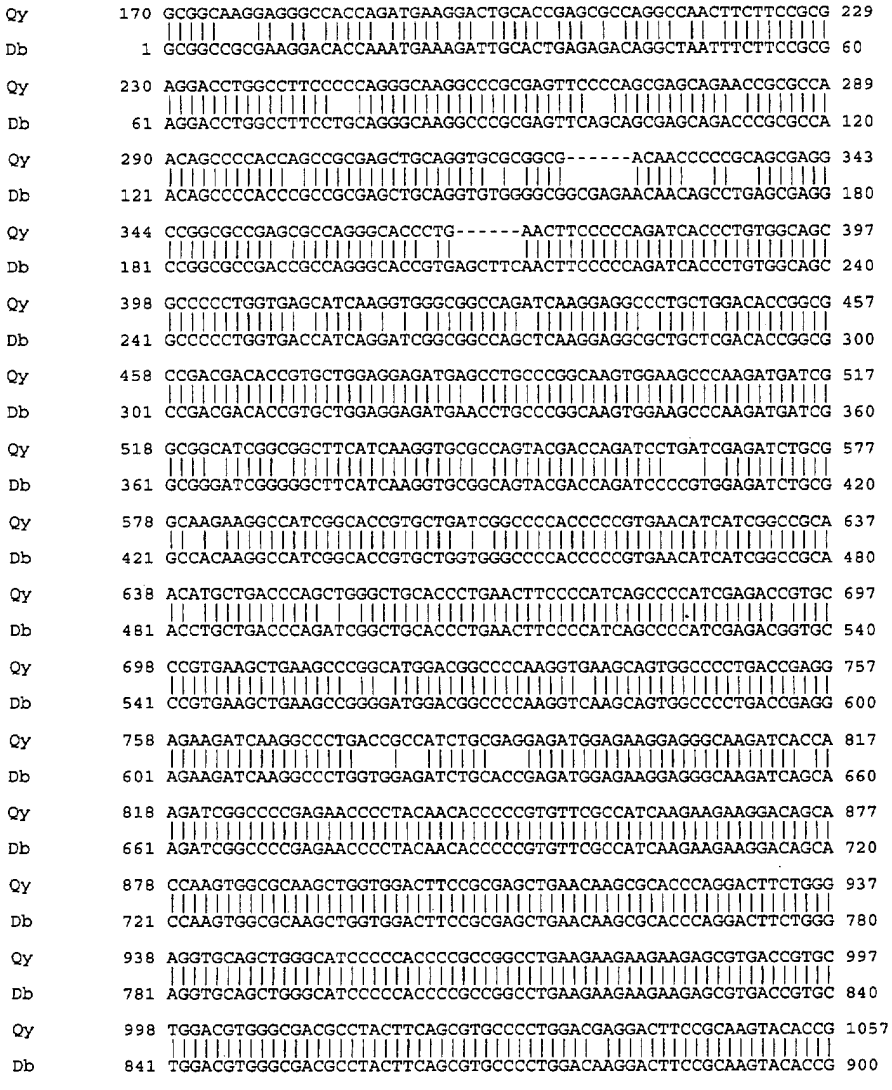


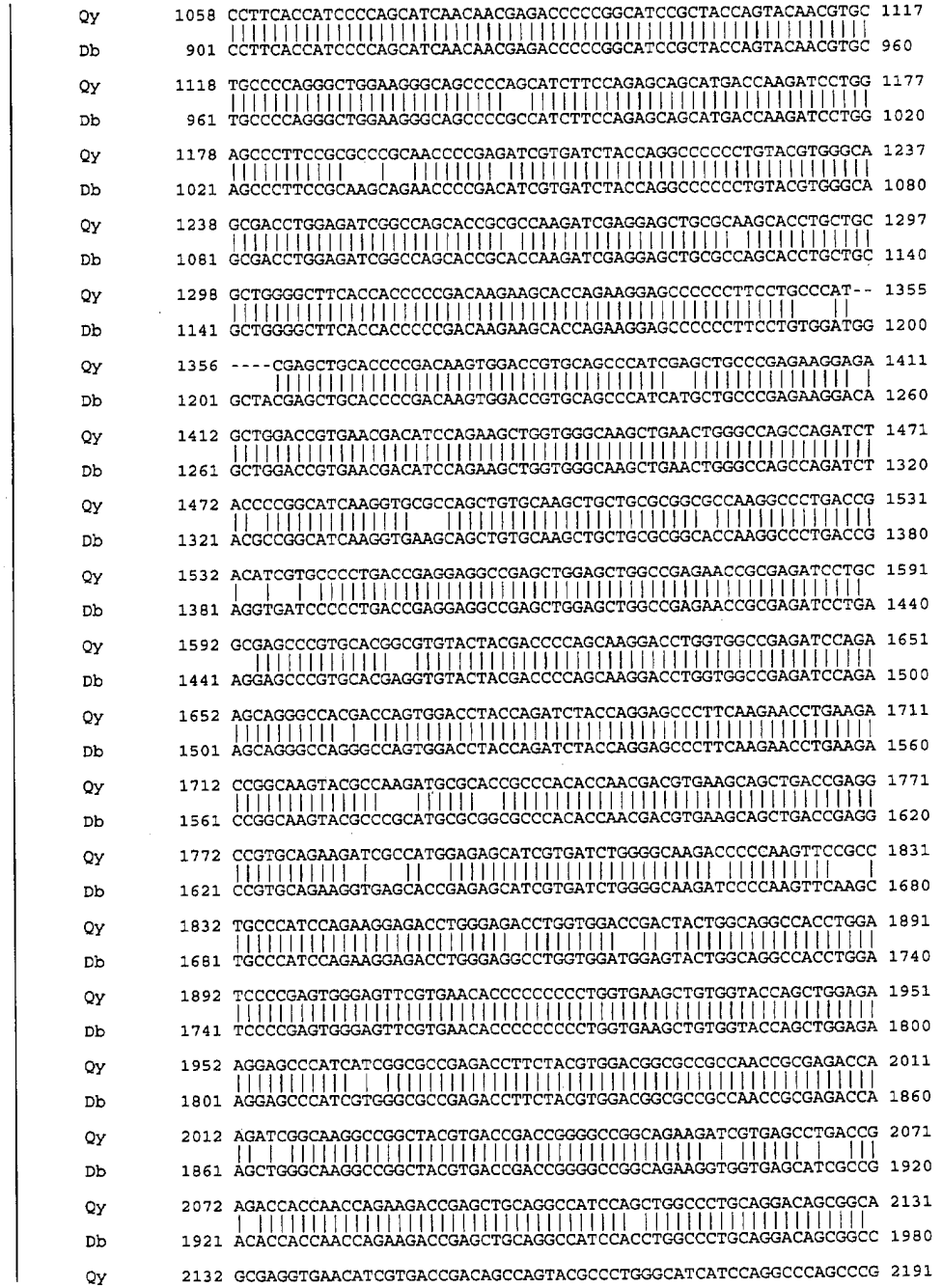


RESULT 7
AR373387
LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 82 from patent US 6602705.
ACCESSION AR373387
VERSION AR373387.1 GI:40075490
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2306)

TITLE Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 82 05-AUG-2003;
FEATURES Location/Qualifiers
source 1. .2306
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 82.2%; Score 2019.2; DB 6; Length 2306;
Best Local Similarity 93.2%; Pred. No. 1.5e-202;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

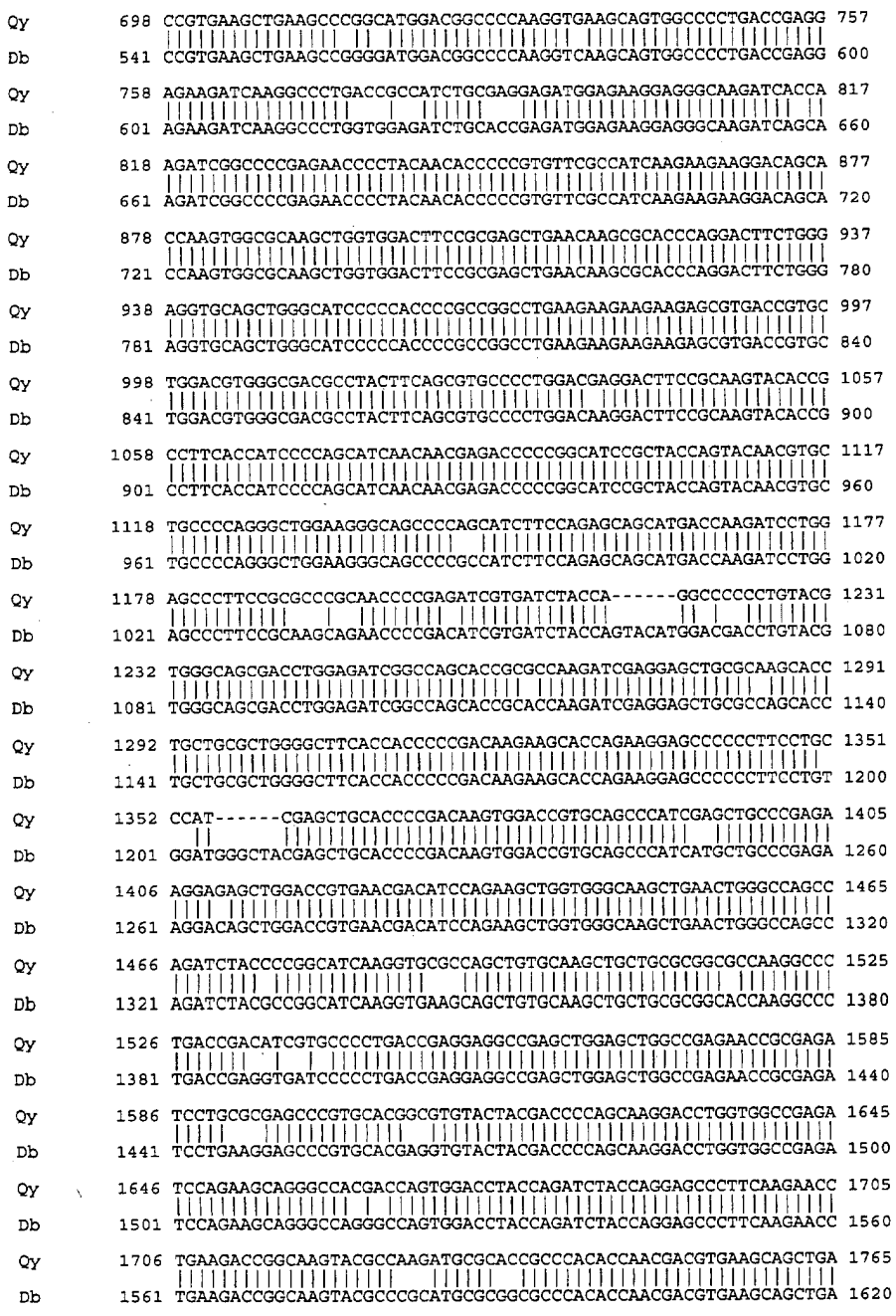


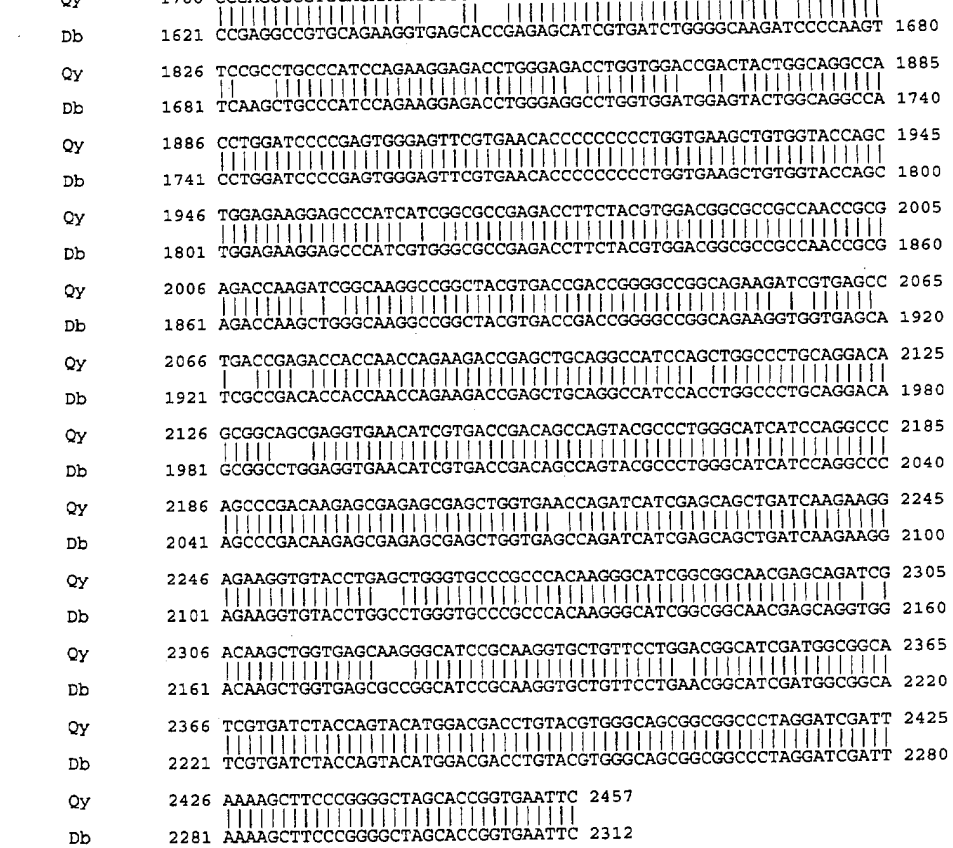


```
Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251

Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311

Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371

Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431

Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457

Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306


RESULT 8
AX427930
LOCUS       AX427930               9166 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 168 from Patent WO0232943.
ACCESSION   AX427930
VERSION     AX427930.1  GI:21538017
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 168 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9166
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             81.4%;  Score 2000.8;  DB 6;  Length 9166;
  Best Local Similarity   91.3%;  Pred. No. 8.8e-201;
  Matches 2185;  Conservative    0;  Mismatches  182;  Indels   25;  Gaps    5;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364

Db       3260 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3319

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421

Db       3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3379

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481

Db       3380 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3439

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541

Db       3440 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3499

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601

Db       3500 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3559

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661

Db       3560 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3619

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721

Db       3620 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3679

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781

Db       3680 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3739

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841

Db       3740 TCTGCACCGAGATGGAGAAGGAGGGCAAGAICAGCAAGATCGGCCCCGAGAACCCCTACA 3799

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901

Db       3800 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3859

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961

Db       3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3919

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021

Db       3920 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3979

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081

Db       3980 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4039

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141

Db       4040 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4099

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201

Db       4100 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4159

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255

Db       4160 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4219

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315

Db       4220 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4279

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCG 1369

Db       4280 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4339

Qy       1370 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429
```

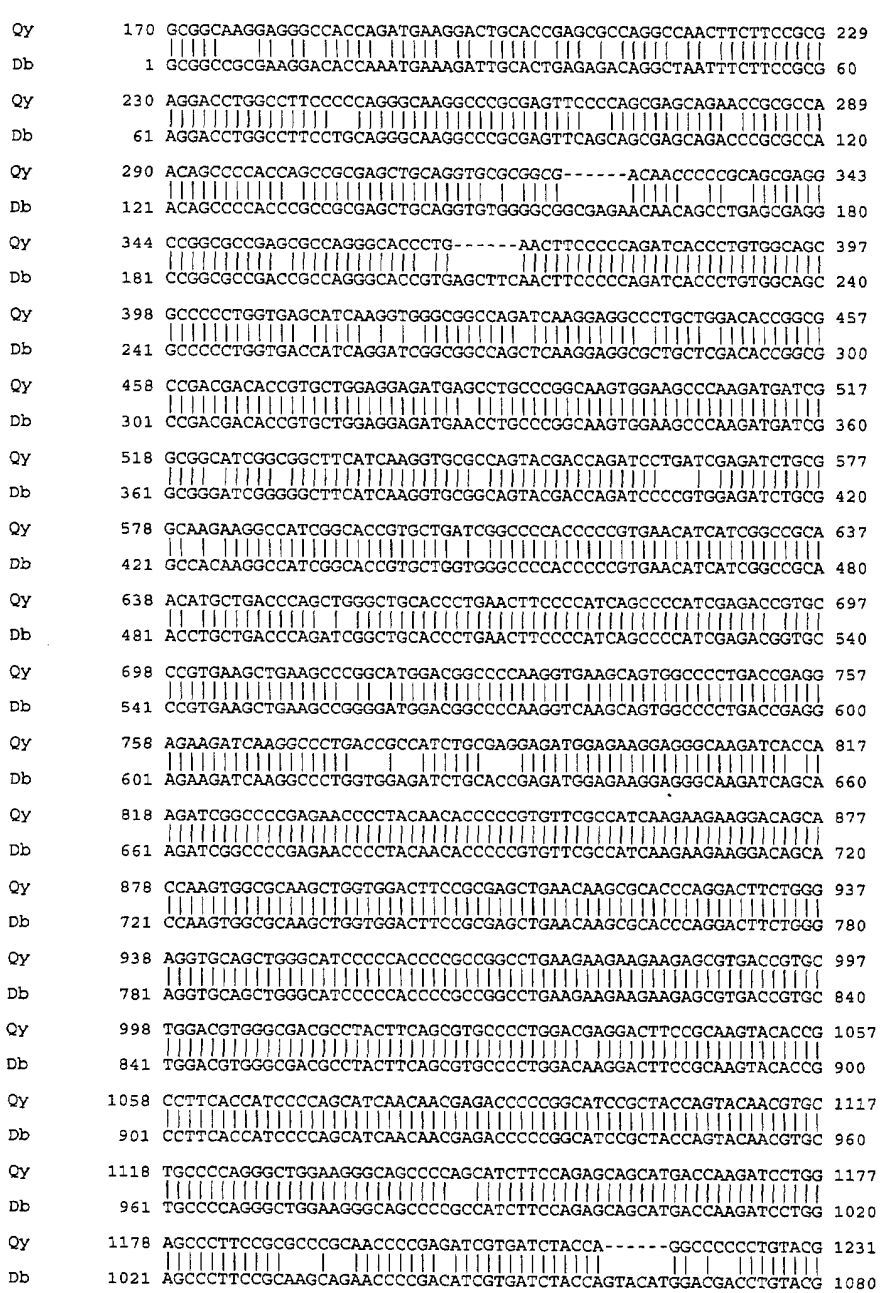
```
Db       4340 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4399

Qy       1430 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1489
Db       4400 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4459

Qy       1490 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1549
Db       4460 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4519

Qy       1550 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1609
Db       4520 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4579

Qy       1610 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1669
Db       4580 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4639

Qy       1670 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1729
Db       4640 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4699

Qy       1730 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1789
Db       4700 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4759

Qy       1790 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849
Db       4760 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4819

Qy       1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909
Db       4820 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4879

Qy       1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969
Db       4880 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4939

Qy       1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
Db       4940 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4999

Qy       2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089
Db       5000 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5059

Qy       2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149
Db       5060 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5119

Qy       2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209
Db       5120 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5179

Qy       2210 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2269
Db       5180 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5239

Qy       2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329
Db       5240 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5299

Qy       2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db       5300 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5351
```

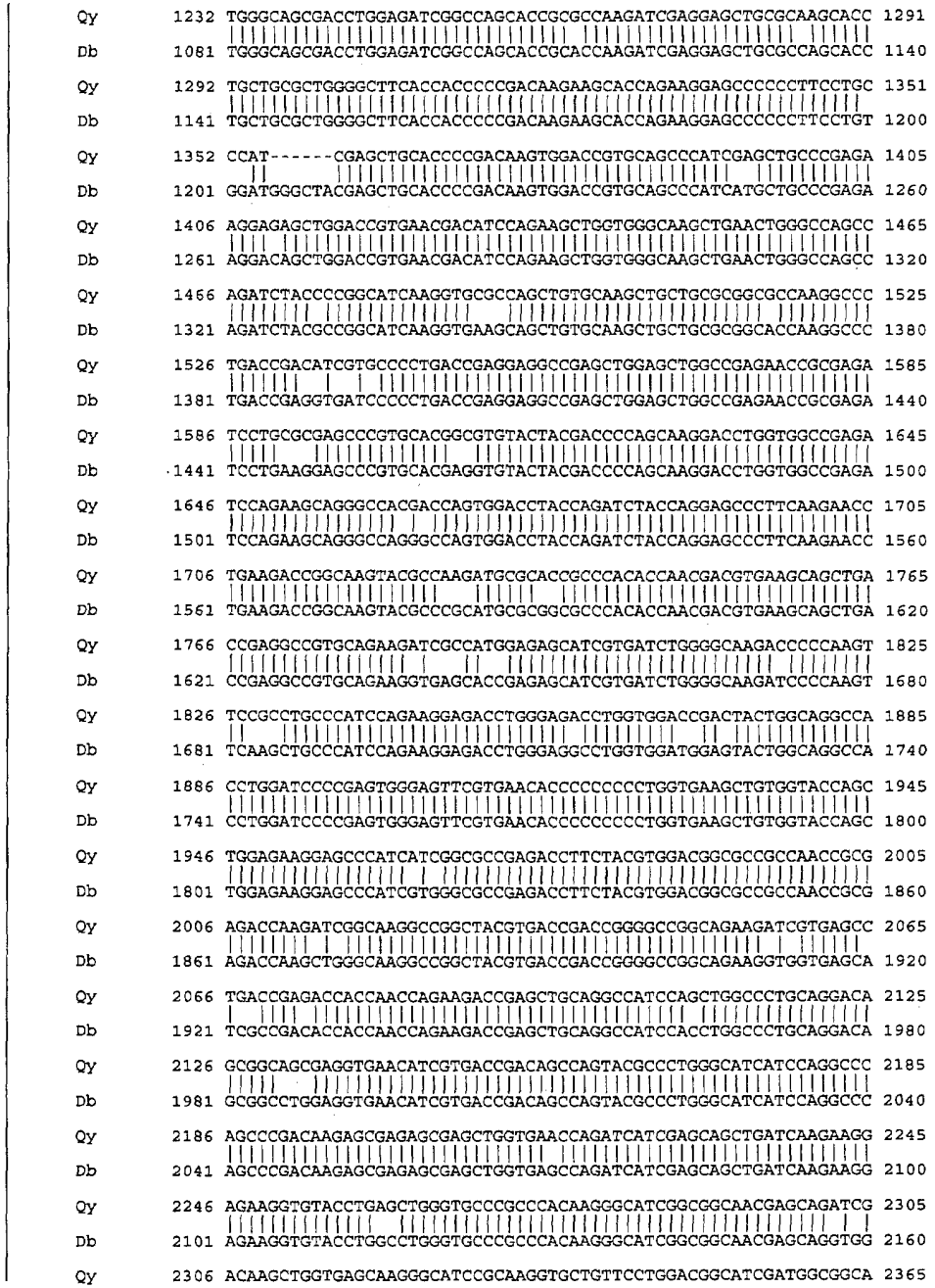
```
RESULT 9
BD263706
LOCUS       BD263706                 2312 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263706
VERSION     BD263706.1  GI:33073474
KEYWORDS    JP 2002533124-A/73.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US     60/114495,01-DEC-1999 US     60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH    Key
                        Location/Qualifiers
            FT   source          1. .2312
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             81.3%;  Score 1998.4;  DB 6;  Length 2312;
  Best Local Similarity   92.9%;  Pred. No. 2.3e-200;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540
```





```
RESULT 10
AR373389
LOCUS       AR373389               2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           81.3%;  Score 1998.4;  DB 6;  Length 2312;
  Best Local Similarity   92.9%;  Pred. No. 2.3e-200;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;
```

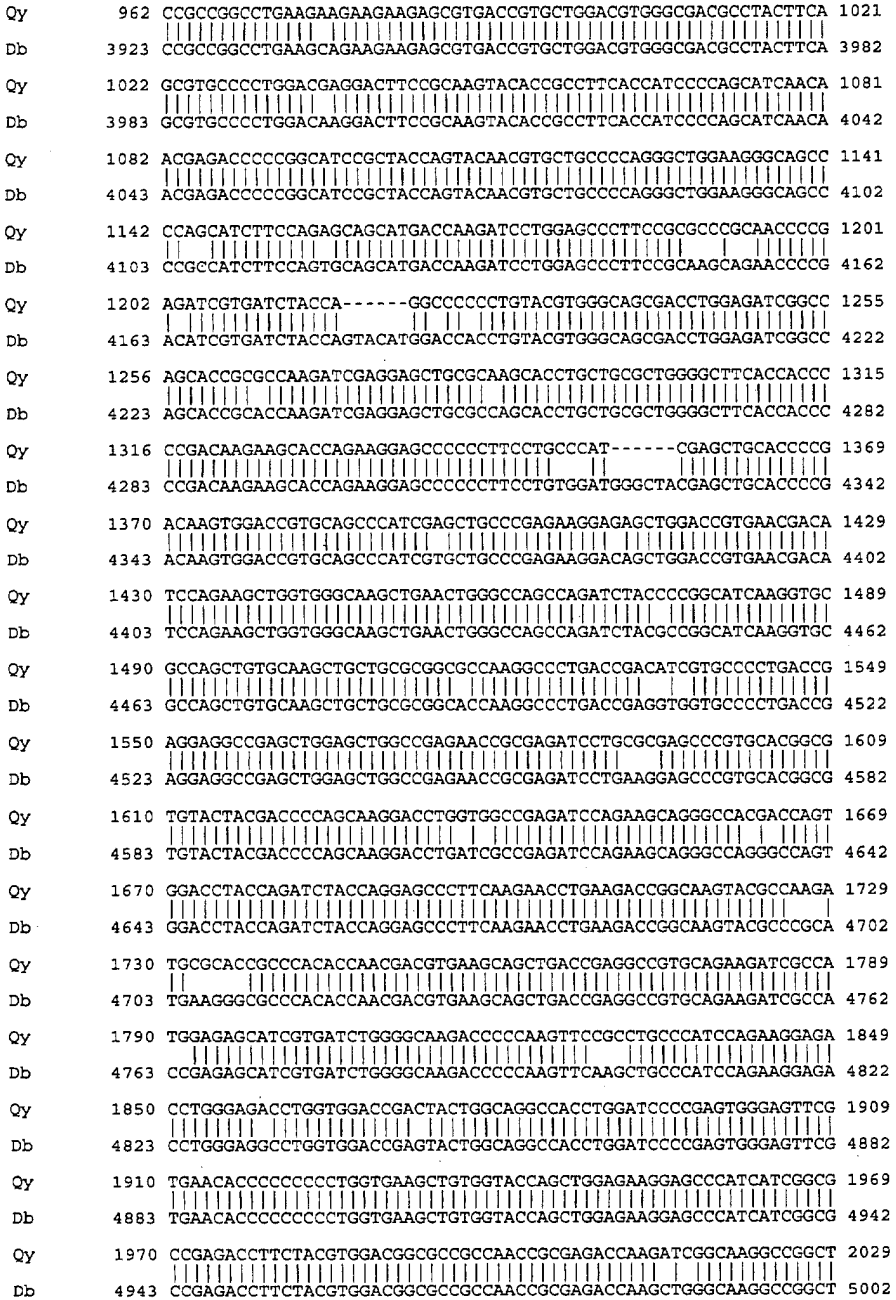


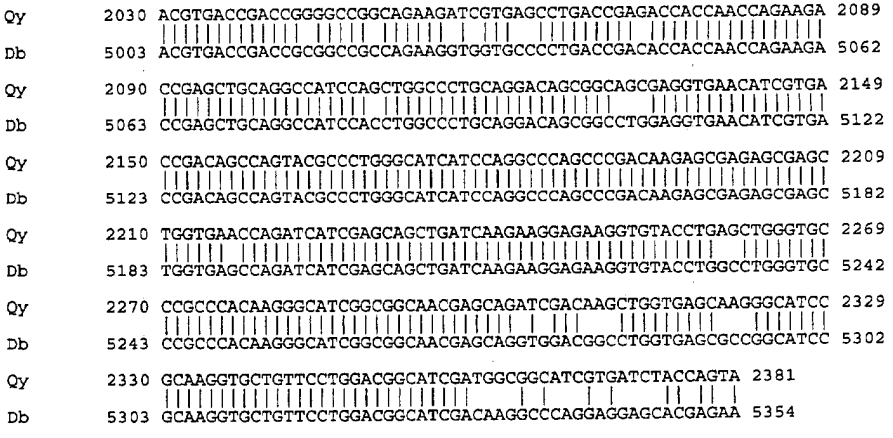


```
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 11
AX427936
LOCUS       AX427936                9788 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             81.1%;  Score 1992.4;  DB 6;  Length 9788;
  Best Local Similarity   91.2%;  Pred. No. 6.5e-200;
  Matches 2181;  Conservative    0;  Mismatches  181;  Indels   30;  Gaps    5;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       3141 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3194

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       3195 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3254

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364

Db       3255 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3314

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421

Db       3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481

Db       3375 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3434

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541

Db       3435 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3494

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601

Db       3495 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3554

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661

Db       3555 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3614

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCCGTGAAGCTGAAGCCCGGCATGG 721

Db       3615 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3674

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781

Db       3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3734

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841

Db       3735 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3794

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901

Db       3795 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3854

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961

Db       3855 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3914

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021

Db       3915 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3974

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081

Db       3975 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4034

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141

Db       4035 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4094

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201

Db       4095 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255

Db       4155 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4214

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315

Db       4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4274

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCG 1369

Db       4275 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4334

Qy       1370 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429

Db       4335 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4394

Qy       1430 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1489

Db       4395 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4454

Qy       1490 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1549

Db       4455 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4514

Qy       1550 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1609
```

```
Db       4515 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4574

Qy       1610 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1669

Db       4575 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4634

Qy       1670 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1729

Db       4635 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4694

Qy       1730 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1789

Db       4695 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4754

Qy       1790 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849

Db       4755 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4814

Qy       1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909

Db       4815 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4874

Qy       1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969

Db       4875 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4934

Qy       1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029

Db       4935 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4994

Qy       2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089

Db       4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054

Qy       2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149

Db       5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114

Qy       2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209

Db       5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

Qy       2210 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2269

Db       5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234

Qy       2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329

Db       5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy       2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381

Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346
```

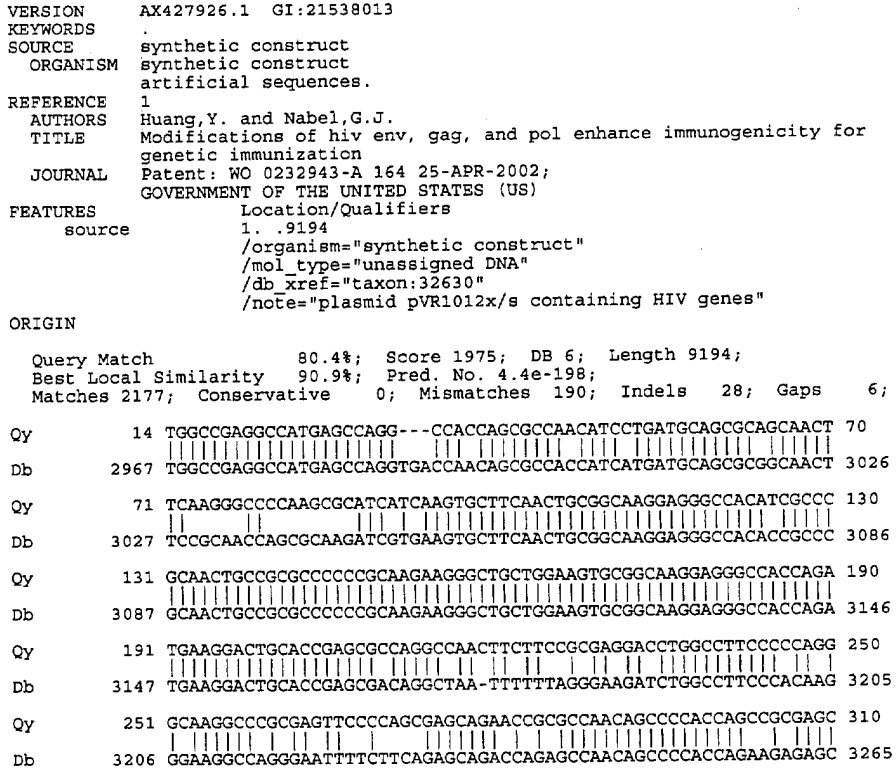
```
RESULT 12
AX427931
LOCUS       AX427931               9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             80.5%;  Score 1978.4;  DB 6;  Length 9169;
  Best Local Similarity   91.1%;  Pred. No. 1.9e-198;
  Matches 2178;  Conservative    0;  Mismatches  186;  Indels   28;  Gaps    6;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364

Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421

Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481

Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541

Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601

Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCA 661

Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721

Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781

Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841

Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901

Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961

Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922
```

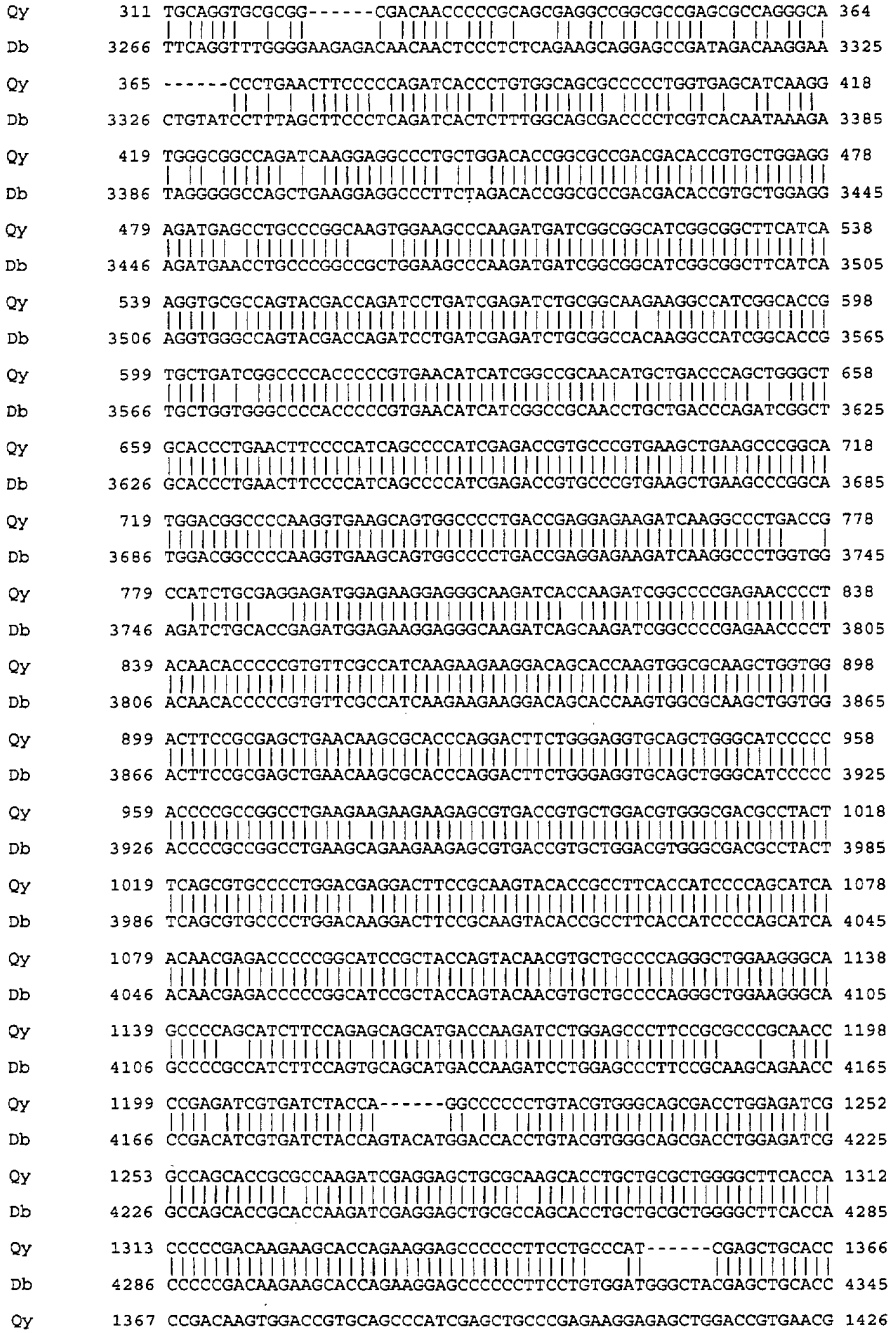


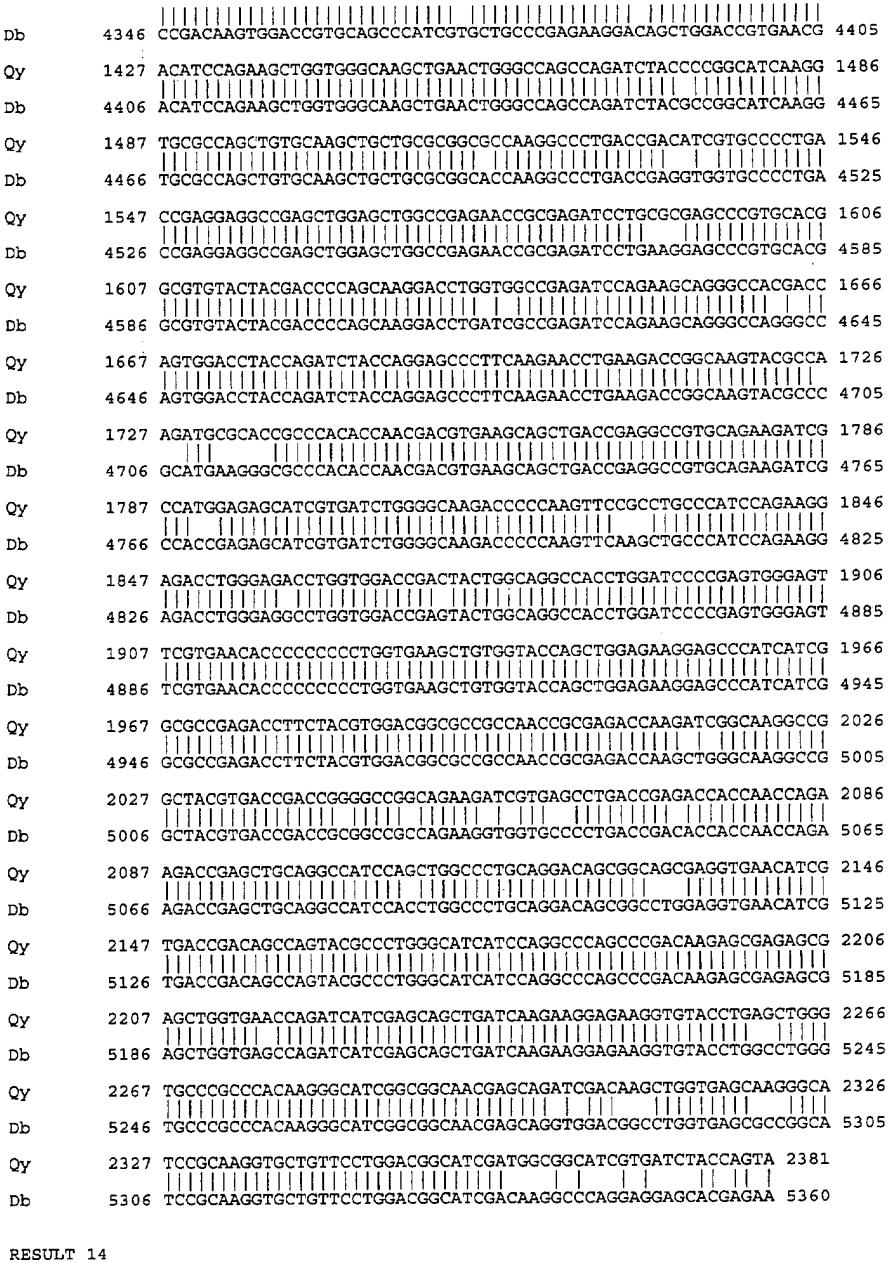


RESULT 13
AX427926
LOCUS AX427926 9194 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 164 from Patent WO0232943.
ACCESSION AX427926
VERSION AX427926.1 GI:21538013
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 164 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES Location/Qualifiers
source 1. .9194
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

Query Match 80.4%; Score 1975; DB 6; Length 9194;
Best Local Similarity 90.9%; Pred. No. 4.4e-198;
Matches 2177; Conservative 0; Mismatches 190; Indels 28; Gaps 6;

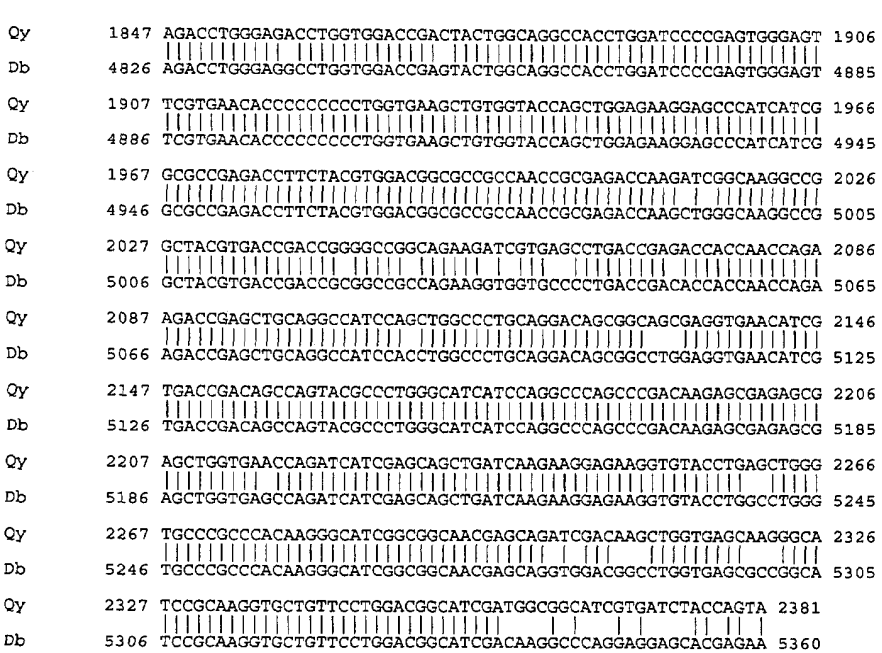






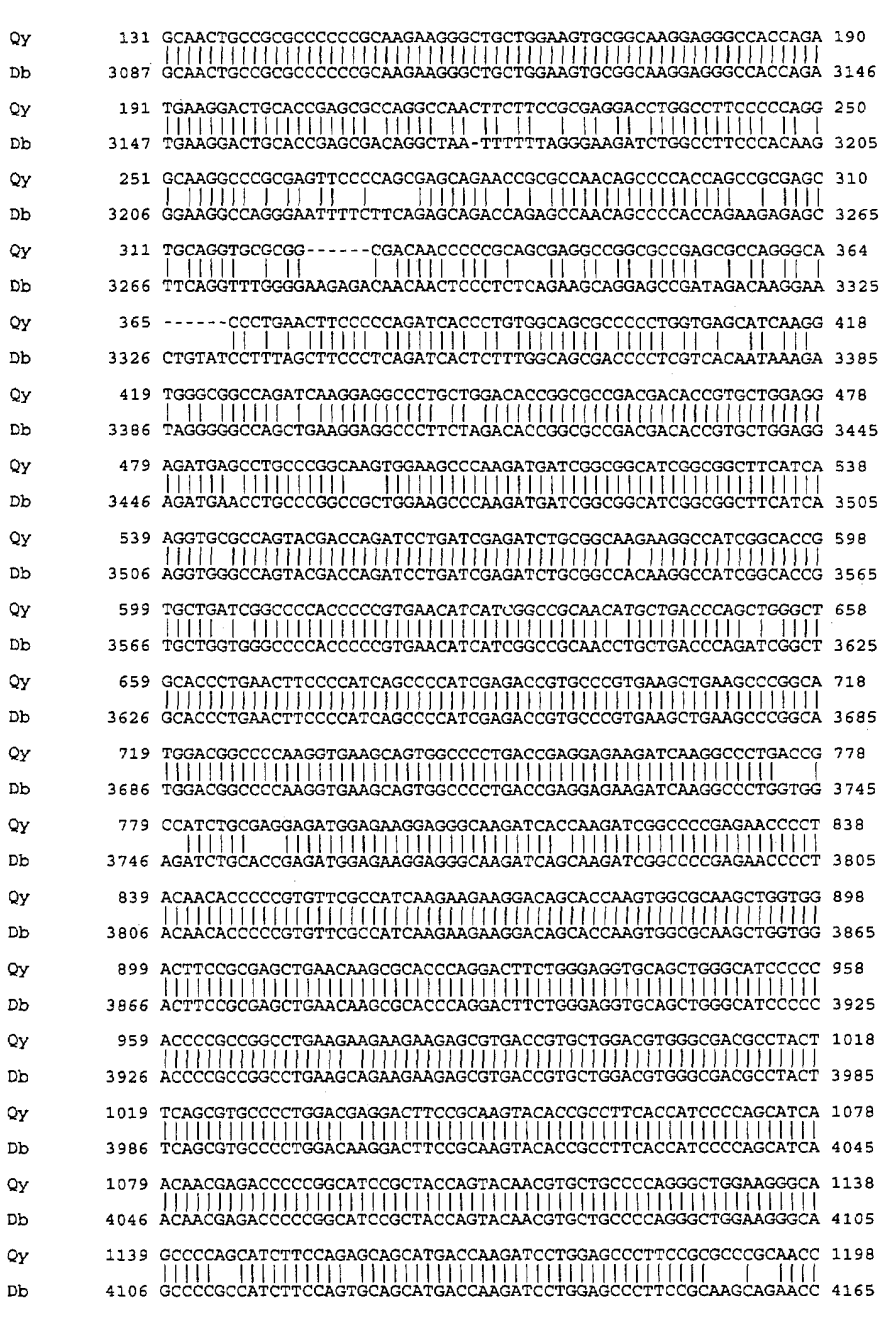
RESULT 14
AX427925
LOCUS AX427925 9194 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 163 from Patent WO0232943.
ACCESSION AX427925

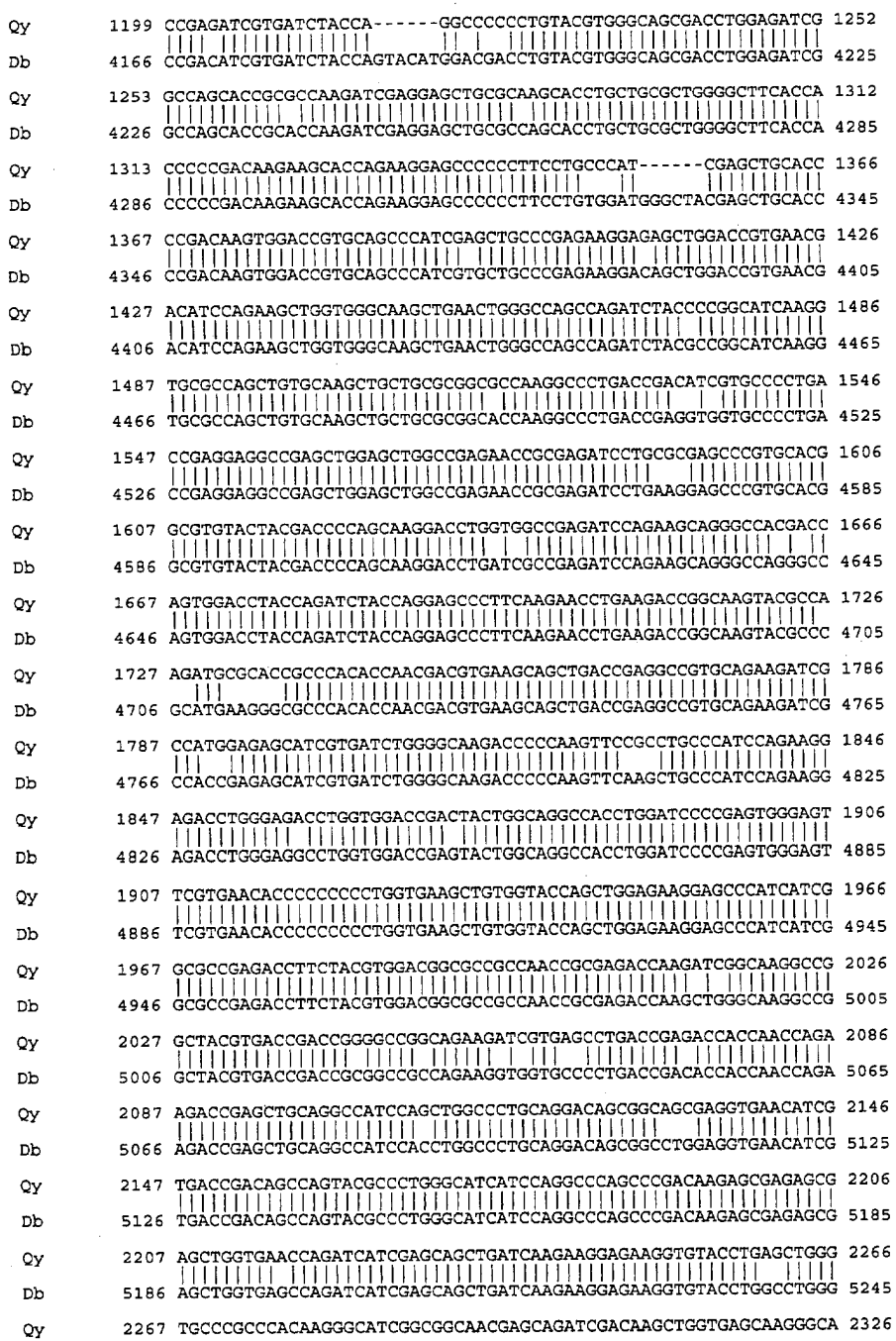
```
VERSION     AX427925.1  GI:21538012
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 163 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             80.3%;  Score 1973.4;  DB 6;  Length 9194;
  Best Local Similarity   90.9%;  Pred. No. 6.5e-198;
  Matches 2176;  Conservative    0;  Mismatches  191;  Indels   28;  Gaps    6;

Qy         14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
Db       2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db       3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db       3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db       3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy        599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db       3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy        659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db       3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy        719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db       3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745

Qy        779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db       3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy        839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db       3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy        899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db       3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy        959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db       3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db       3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db       4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db       4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy       1199 CCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCG 1252
Db       4166 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy       1253 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1312
Db       4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy       1313 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACC 1366
Db       4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy       1367 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1426
Db       4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405

Qy       1427 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1486
Db       4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy       1487 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1546
Db       4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy       1547 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1606
Db       4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy       1607 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1666
Db       4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645

Qy       1667 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1726
Db       4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy       1727 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1786
Db       4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy       1787 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1846
Db       4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825
```



RESULT 15
AX427927
LOCUS AX427927 12411 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 165 from Patent WO0232943.
ACCESSION AX427927
VERSION AX427927.1 GI:21538014
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 165 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES Location/Qualifiers
source 1. .12411
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

Query Match 80.3%; Score 1973.4; DB 6; Length 12411;
Best Local Similarity 90.9%; Pred. No. 6e-198;
Matches 2176; Conservative 0; Mismatches 191; Indels 28; Gaps 6;





Search completed: April 10, 2004, 12:49:25
Job time : 6322.25 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 12:49:38 ; Search time 591.888 Seconds
(without alignments)
15572.780 Million cell updates/sec

Title: US-09-610-313-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2475585 seqs, 1875730760 residues

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2443.4	99.4	2445	14	US-10-190-435-43	Sequence 43, Appl
2	2443.4	99.4	2445	14	US-10-190-305A-37	Sequence 37, Appl
3	2436.4	99.2	3930	14	US-10-190-435-9	Sequence 9, Appli
4	2434.8	99.1	3930	14	US-10-190-435-10	Sequence 10, Appl
5	2434.8	99.1	3930	14	US-10-190-435-11	Sequence 11, Appl
6	2434.8	99.1	5184	14	US-10-190-435-58	Sequence 58, Appl
7	2434.8	99.1	5184	14	US-10-190-305A-82	Sequence 82, Appl
8	2428.6	98.8	2457	14	US-10-190-435-44	Sequence 44, Appl
9	2428.6	98.8	2457	14	US-10-190-305A-38	Sequence 38, Appl
10	2404.4	97.9	3531	14	US-10-190-435-13	Sequence 13, Appl
11	2403.4	97.8	2457	14	US-10-190-435-45	Sequence 45, Appl
12	2403.4	97.8	2457	14	US-10-190-305A-39	Sequence 39, Appl
13	2402.8	97.8	3537	14	US-10-190-435-14	Sequence 14, Appl
14	2402.8	97.8	3537	14	US-10-190-435-15	Sequence 15, Appl
15	2401.8	97.8	5145	14	US-10-190-435-12	Sequence 12, Appl
16	2401.8	97.8	5145	14	US-10-190-305A-12	Sequence 12, Appl
17	2343.4	95.4	3607	14	US-10-190-435-48	Sequence 48, Appl
18	2343.4	95.4	3607	14	US-10-190-305A-42	Sequence 42, Appl
19	2325.2	94.6	3597	14	US-10-190-435-46	Sequence 46, Appl
20	2325.2	94.6	3597	14	US-10-190-305A-40	Sequence 40, Appl
21	2295.8	93.4	3624	14	US-10-190-435-47	Sequence 47, Appl
22	2295.8	93.4	3624	14	US-10-190-305A-41	Sequence 41, Appl
23	2136	86.9	2460	14	US-10-241-009-30	Sequence 30, Appl
24	2136	86.9	2460	14	US-10-190-434B-30	Sequence 30, Appl
25	2136	86.9	2460	14	US-10-190-305A-32	Sequence 32, Appl
26	2135	86.9	3564	14	US-10-241-009-13	Sequence 13, Appl
27	2135	86.9	3564	14	US-10-241-009-14	Sequence 14, Appl
28	2135	86.9	3564	14	US-10-190-434B-13	Sequence 13, Appl
29	2135	86.9	3564	14	US-10-190-434B-14	Sequence 14, Appl
30	2134.4	86.9	4716	14	US-10-190-435-17	Sequence 17, Appl
31	2134.4	86.9	4716	14	US-10-190-305A-13	Sequence 13, Appl
32	2130.6	86.7	3999	14	US-10-241-009-9	Sequence 9, Appli
33	2130.6	86.7	3999	14	US-10-190-434B-9	Sequence 9, Appli
34	2129	86.7	3999	14	US-10-241-009-10	Sequence 10, Appl
35	2129	86.7	3999	14	US-10-241-009-11	Sequence 11, Appl
36	2129	86.7	3999	14	US-10-190-434B-10	Sequence 10, Appl
37	2129	86.7	3999	14	US-10-190-434B-11	Sequence 11, Appl
38	2129	86.7	5283	14	US-10-241-009-54	Sequence 54, Appl
39	2129	86.7	5283	14	US-10-190-434B-54	Sequence 54, Appl
40	2129	86.7	5283	14	US-10-190-305A-75	Sequence 75, Appl
41	2128.8	86.6	4713	14	US-10-190-435-59	Sequence 59, Appl
42	2128.8	86.6	4713	14	US-10-190-305A-83	Sequence 83, Appl
43	2128.4	86.6	3462	14	US-10-190-435-16	Sequence 16, Appl
44	2115.2	86.1	2466	14	US-10-241-009-31	Sequence 31, Appl
45	2115.2	86.1	2466	14	US-10-190-434B-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMWM_C
US-10-190-435-43

Query Match 99.4%; Score 2443.4; DB 14; Length 2445;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
```

```
Db      121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180
Qy      187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db      181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240
Qy      247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db      241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300
Qy      307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db      301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360
Qy      367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db      361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420
Qy      427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db      421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480
Qy      487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db      481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540
Qy      547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db      541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600
Qy      607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db      601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660
Qy      667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db      661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720
Qy      727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db      721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780
Qy      787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db      781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840
Qy      847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db      841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy      907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db      901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy      967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db      961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy     1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db     1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy     1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db     1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy     1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db     1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy     1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
```

```
Db     1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260
Qy     1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db     1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320
Qy     1327 CACCAGAAGGAGCCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
Db     1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380
Qy     1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
Db     1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440
Qy     1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
Db     1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500
Qy     1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
Db     1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560
Qy     1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
Db     1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620
Qy     1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
Db     1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680
Qy     1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
Db     1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740
Qy     1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
Db     1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800
Qy     1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
Db     1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860
Qy     1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
Db     1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920
Qy     1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
Db     1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980
Qy     1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
Db     1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040
Qy     2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
Db     2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100
Qy     2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
Db     2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160
Qy     2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
Db     2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220
Qy     2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
Db     2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280
Qy     2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
Db     2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340
```

```
Qy       2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy       2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445


RESULT 2
US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: p2Pol.opt.YMWM_C
US-10-190-305A-37

  Query Match             99.4%;  Score 2443.4;  DB 14;  Length 2445;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2444;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380

Qy       1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440

Qy       1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500

Qy       1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560

Qy       1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620
```

```
Qy       1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680

Qy       1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740

Qy       1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800

Qy       1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860

Qy       1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920

Qy       1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980

Qy       1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040

Qy       2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100

Qy       2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160

Qy       2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220

Qy       2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280

Qy       2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340

Qy       2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy       2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 3
US-10-190-435-9
; Sequence 9, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut_C
US-10-190-435-9

Query Match 99.2%; Score 2436.4; DB 14; Length 3930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

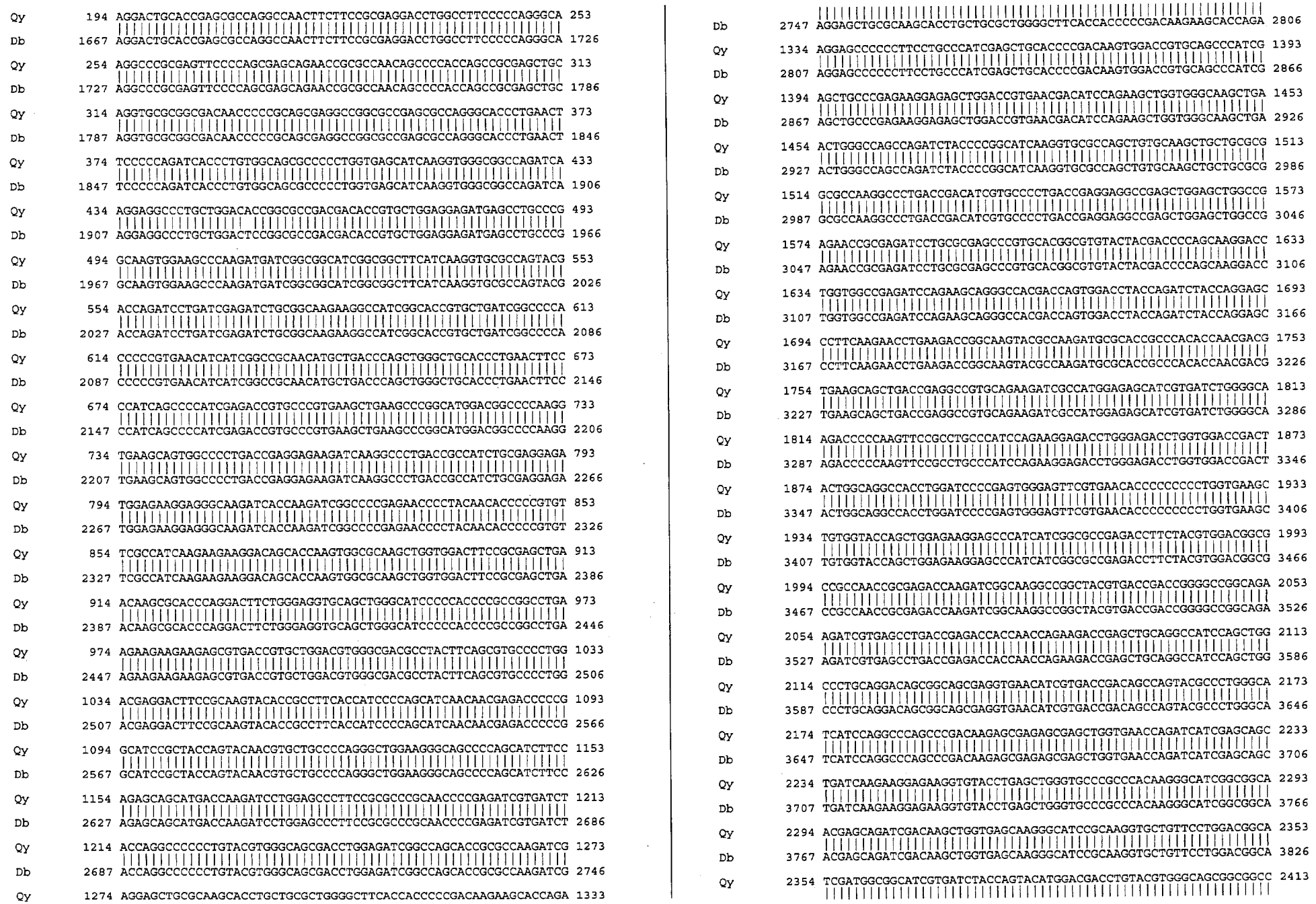
```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
```

```
Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy      1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db      2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy      1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db      2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy      1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db      2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy      1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db      2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy      1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db      3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy      1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db      3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy      1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db      3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy      1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db      3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy      1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db      3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy      1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db      3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy      1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db      3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy      1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db      3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy      2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db      3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy      2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db      3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy      2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db      3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy      2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db      3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy      2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db      3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy      2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db      3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy      2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db      3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 4
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

Query Match 99.1%; Score 2434.8; DB 14; Length 3930;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db      1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db      1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy       134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db      1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
```


Qy 194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db 1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db 2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746
Qy 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db 2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806
Qy 1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db 2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866
Qy 1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db 2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926
Qy 1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db 2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986
Qy 1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db 2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046
Qy 1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db 3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106
Qy 1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db 3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166
Qy 1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db 3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226
Qy 1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db 3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286
Qy 1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db 3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346
Qy 1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db 3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406
Qy 1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db 3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466
Qy 1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db 3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526
Qy 2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db 3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586
Qy 2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db 3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646
Qy 2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db 3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706
Qy 2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db 3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766
Qy 2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db 3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826
Qy 2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413

```
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             ||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 5
US-10-190-435-11
; Sequence 11, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna_C
US-10-190-435-11

  Query Match             99.1%;  Score 2434.8;  DB 14;  Length 3930;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

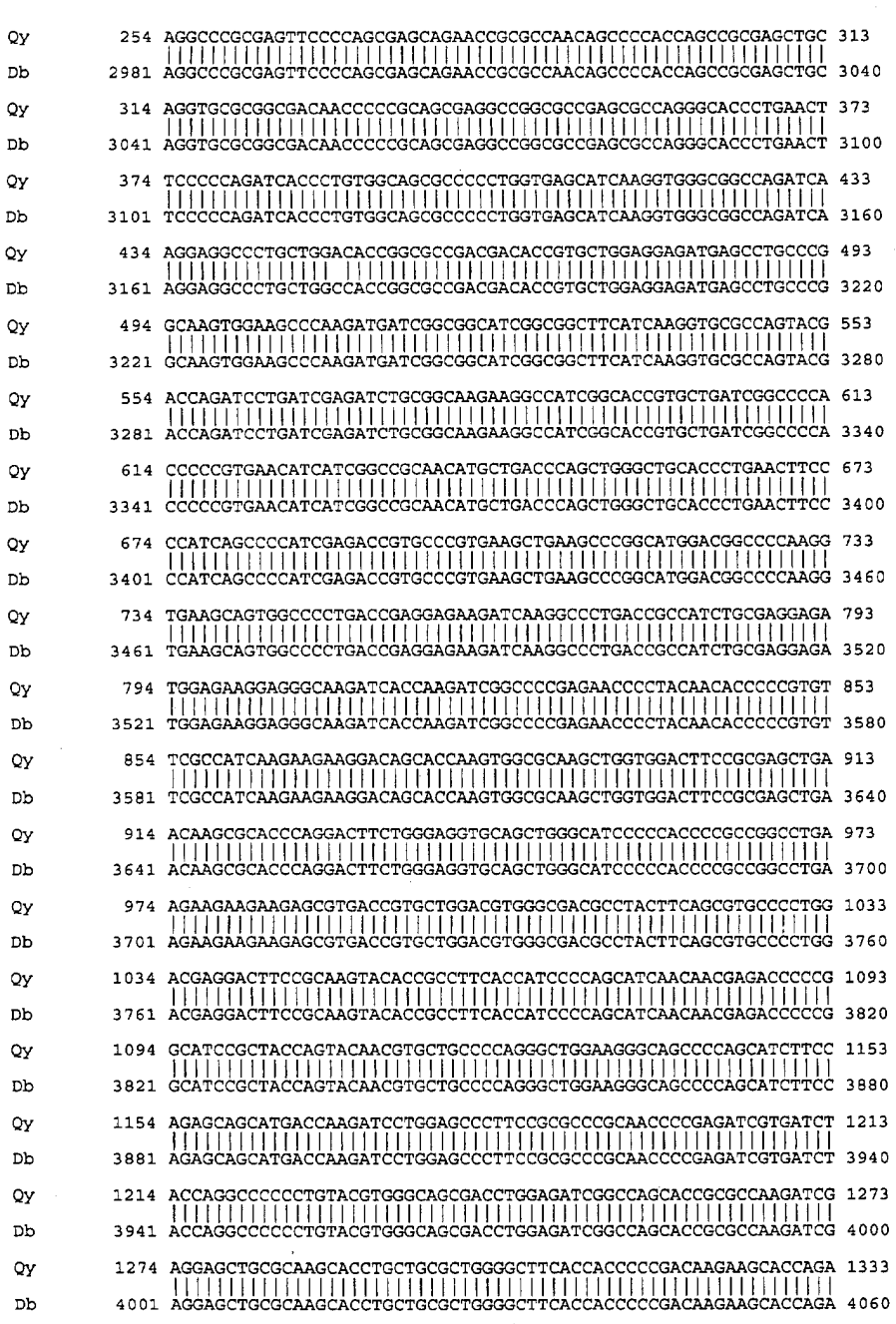
```
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCCCACAAGGGCATCGGCGGCA 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 6
US-10-190-435-58
; Sequence 58, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
;  LENGTH: 5184
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C
US-10-190-435-58

  Query Match             99.1%;  Score 2434.8;  DB 14;  Length 5184;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700
Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880
Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000
Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060
Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db       4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120
Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db       4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180
Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db       4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240
Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db       4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300
Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360
Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420
Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480
Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540
Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600
Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660
Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720
Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780
Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840
Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900
Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960
Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020
Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080
Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140
Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 7
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 5184
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: TatRevNefgagCpolIna C
US-10-190-305A-82

Query Match 99.1%; Score 2434.8; DB 14; Length 5184;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860
Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       2861 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980
```


Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy 1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db 3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000
Qy 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db 4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060
Qy 1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db 4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120
Qy 1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db 4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180
Qy 1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db 4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240
Qy 1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db 4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300
Qy 1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db 4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360
Qy 1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db 4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420
Qy 1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db 4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480
Qy 1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db 4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540
Qy 1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db 4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600
Qy 1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db 4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660
Qy 1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db 4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720
Qy 1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db 4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780
Qy 2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db 4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840
Qy 2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db 4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900
Qy 2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db 4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960
Qy 2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db 4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020
Qy 2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db 5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080
Qy 2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db 5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140
Qy 2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451

```
                 ||||||||||||||||||||||||||||||||||||||||
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 8
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44

Query Match 98.8%; Score 2428.6; DB 14; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              |||||||||||||||||||||||||||||||||||||    ||      ||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
```

```
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451

Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 9
US-10-190-305A-38
; Sequence 38, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C

US-10-190-305A-38

Query Match 98.8%; Score 2428.6; DB 14; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
```

```
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
```

```
Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 10
US-10-190-435-13
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 3531
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13

  Query Match   97.9%; Score 2404.4; DB 14; Length 3531;
  Best Local Similarity   99.1%; Pred. No. 0;
  Matches 2417; Conservative   0; Mismatches   21; Indels   0; Gaps   0;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1208 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387
```

```
Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2048 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2107

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2108 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2167

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2168 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2227

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2288 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2347

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2348 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2407

Qy      1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2408 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2467
```

```
Qy      1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2468 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2527

Qy      1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2528 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2587

Qy      1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2588 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 2647

Qy      1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2648 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 2707

Qy      1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2708 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 2767

Qy      1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2768 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 2827

Qy      1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2828 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 2887

Qy      1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2888 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 2947

Qy      1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2948 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3007

Qy      1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3008 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3067

Qy      1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3068 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3127

Qy      2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3128 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3187

Qy      2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3188 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3247

Qy      2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3248 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3307

Qy      2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3308 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3367

Qy      2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3368 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3427

Qy      2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3428 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3487

Qy      2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
            |||||||||||||||||||||||||||||||||||||
Db      3488 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3525
```

```
RESULT 11
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

  Query Match             97.8%;  Score 2403.4;  DB 14;  Length 2457;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2439;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              |||||||||||      || | ||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              |||||||||||||||||||||||||||||||||||||  ||      ||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
```

```
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 12
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

Query Match 97.8%; Score 2403.4; DB 14; Length 2457;
Best Local Similarity 99.3%; Pred. No. 0;

Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

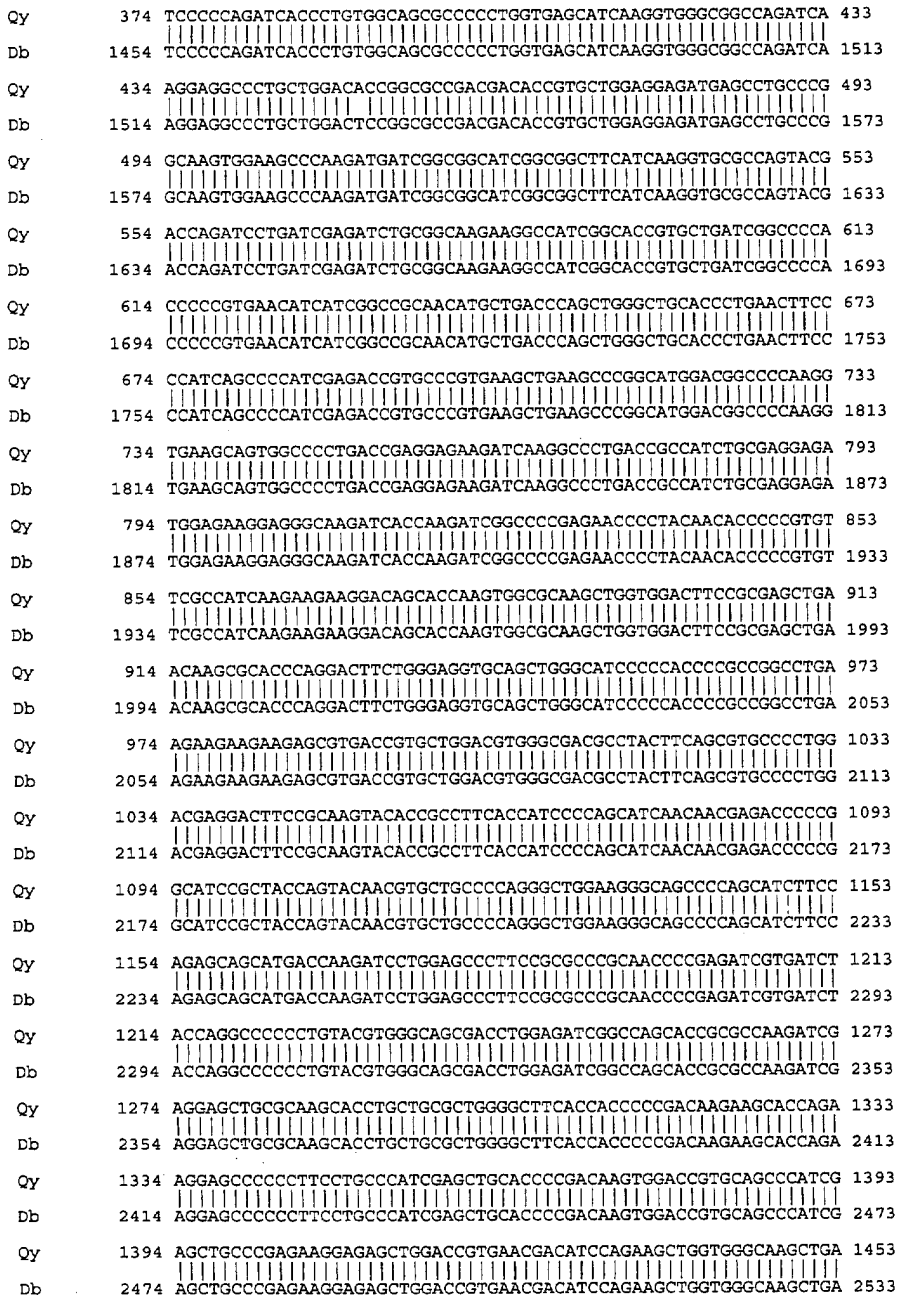
```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60
Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120
Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180
Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240
Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300
Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360
Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420
Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480
Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540
Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600
Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660
Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720
Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780
Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840
Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
```

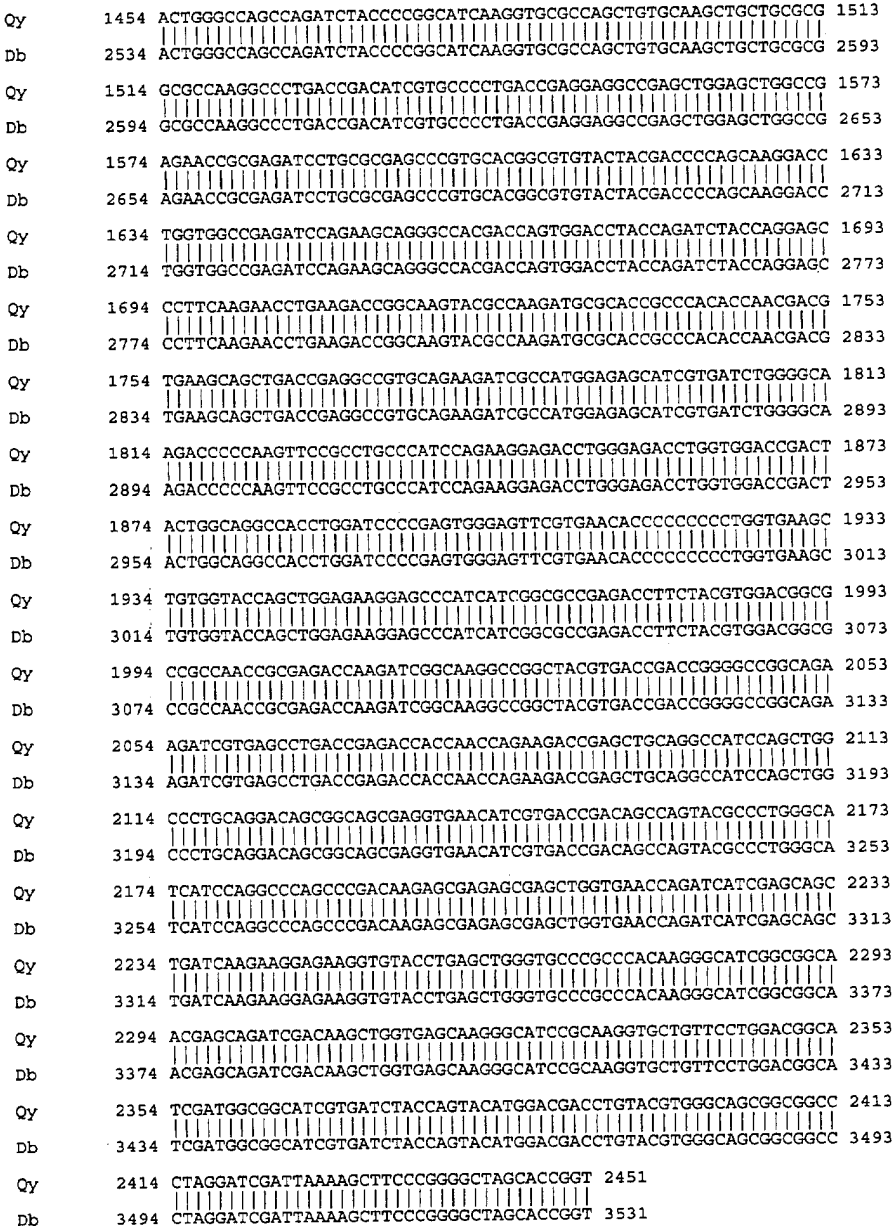
```
Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
             |||||||||||      || |||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
             |||||||||||||||||||||||||||||||||||||    ||      ||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCC 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 13
US-10-190-435-14
; Sequence 14, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 3537
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmutAtt_C
US-10-190-435-14

Query Match 97.8%; Score 2402.8; DB 14; Length 3537;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2416; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| |||||||||||| || |  |  ||  ||  | || ||| |      |
Db       1094 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1153

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1154 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1213

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1214 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1273

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1274 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1333

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1334 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1393

Qy        314 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1394 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1453
```





RESULT 14
US-10-190-435-15
; Sequence 15, Application US/10190435
; Publication No. US20030143248A1

; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 3537
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmutIna_C
US-10-190-435-15

Query Match 97.8%; Score 2402.8; DB 14; Length 3537;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2416; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy      14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
           |||||||||| |||||||||||| || |  |  ||  ||  | || ||| | |     |
Db    1094 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1153

Qy      74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1154 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1213

Qy     134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1214 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1273

Qy     194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1274 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1333

Qy     254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1334 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1393

Qy     314 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1394 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1453

Qy     374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1454 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1513

Qy     434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1514 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1573

Qy     494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1574 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1633

Qy     554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1634 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1693

Qy     614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1694 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1753

Qy     674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1754 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1813

Qy     734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1814 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1873

Qy     794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1874 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1933

Qy     854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1934 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1993

Qy     914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1994 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2053

Qy     974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2054 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2113

Qy    1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2114 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2173

Qy    1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2174 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2233

Qy    1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2234 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2293

Qy    1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2294 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2353

Qy    1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2354 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2413

Qy    1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2414 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2473

Qy    1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2474 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2533

Qy    1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2534 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2593

Qy    1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2594 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 2653

Qy    1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2654 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 2713

Qy    1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2714 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 2773

Qy    1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2774 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 2833

Qy    1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2834 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 2893

Qy    1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
```

```
Db       2894 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 2953
Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db       2954 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3013
Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db       3014 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3073
Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db       3074 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3133
Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db       3134 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3193
Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db       3194 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3253
Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db       3254 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3313
Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db       3314 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3373
Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db       3374 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3433
Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db       3434 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3493
Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       3494 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3531
```

RESULT 15
US-10-190-435-12
; Sequence 12, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagComplPolmutInaTatRevNef_C
US-10-190-435-12

Query Match 97.8%; Score 2401.8; DB 14; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2403; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606
Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
```

```
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGG 2418
             |||||
Db       3887 CTAGG 3891
```

Search completed: April 10, 2004, 21:24:47
Job time : 597.888 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:48:32 ; Search time 118.045 Seconds
(without alignments)
11550.795 Million cell updates/sec

Title: US-09-610-313-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

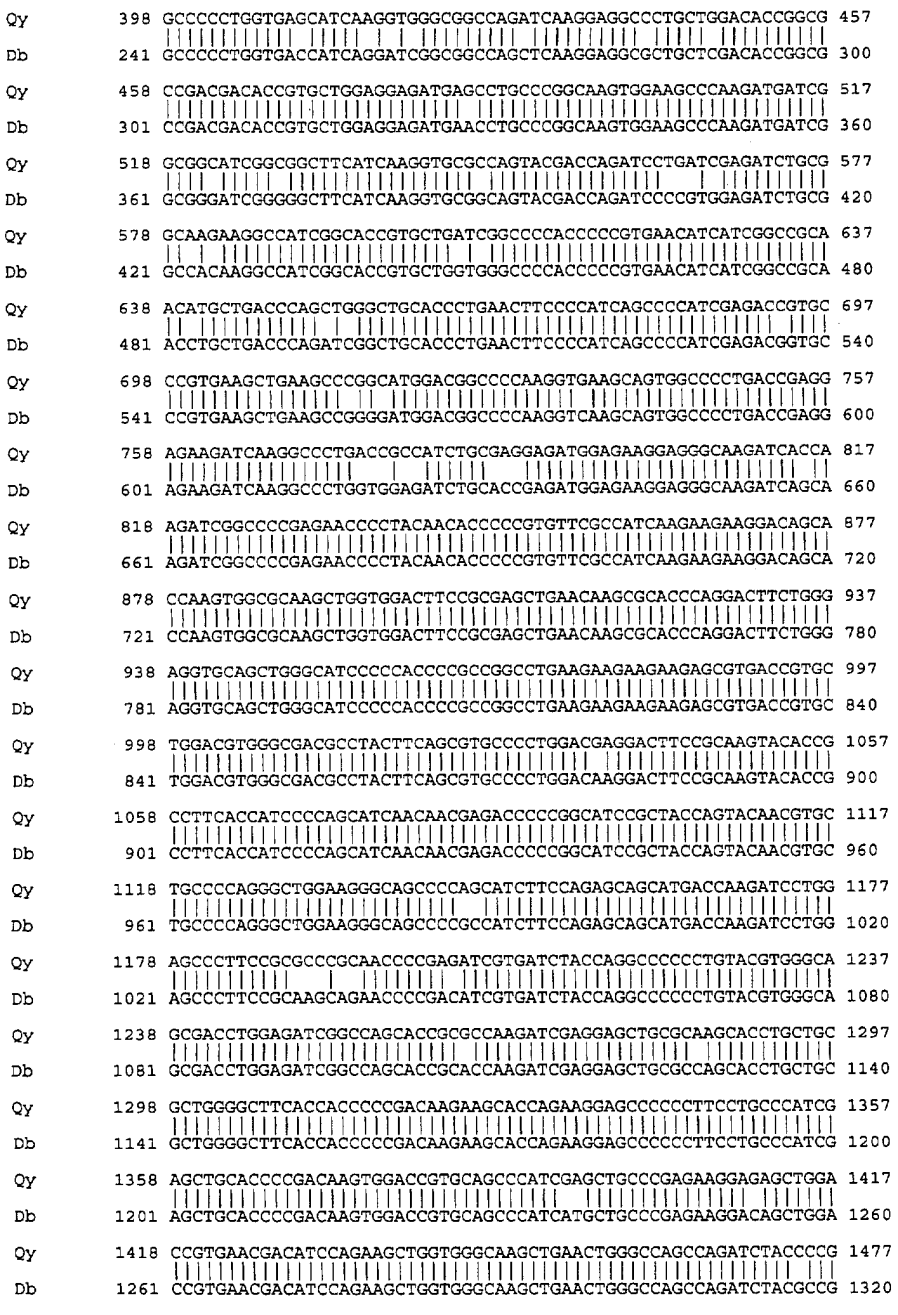
Result No.	Score	% Query Match	Length	DB	ID	Description
1	2040	83.0	2300	4	US-09-475-515-83	Sequence 83, Appl
2	2019.2	82.2	2306	4	US-09-475-515-82	Sequence 82, Appl
3	1998.4	81.3	2312	4	US-09-475-515-84	Sequence 84, Appl
4	1915.8	78.0	4319	4	US-09-475-515-6	Sequence 6, Appli
5	1872.8	76.2	2299	4	US-09-475-515-81	Sequence 81, Appl
6	1852	75.4	2305	4	US-09-475-515-80	Sequence 80, Appl
7	1626.6	66.2	4307	4	US-09-552-950-2	Sequence 2, Appli
8	1599.4	65.1	9772	4	US-09-552-950-5	Sequence 5, Appli
9	1539.2	62.6	8366	4	US-09-872-733A-6	Sequence 6, Appli
10	1503.8	61.2	4338	4	US-09-872-733A-1	Sequence 1, Appli
11	1178.4	48.0	9010	4	US-09-184-418C-8	Sequence 8, Appli
12	1147.4	46.7	8972	4	US-09-184-418C-9	Sequence 9, Appli
13	1140.2	46.4	8959	4	US-09-184-418C-11	Sequence 11, Appl
14	1137	46.3	2467	4	US-09-872-733A-3	Sequence 3, Appli
15	1116.8	45.5	8992	4	US-09-184-418C-4	Sequence 4, Appli
16	1081.6	44.0	2601	3	US-09-117-217-7	Sequence 7, Appli
17	1081.6	44.0	2601	3	US-09-117-217-9	Sequence 9, Appli
18	1081.6	44.0	2601	3	US-09-117-217-11	Sequence 11, Appl
19	1081.6	44.0	2601	3	US-09-117-217-13	Sequence 13, Appl
20	1081.6	44.0	2601	4	US-09-735-487-7	Sequence 7, Appli
21	1081.6	44.0	2601	4	US-09-735-487-9	Sequence 9, Appli
22	1081.6	44.0	2601	4	US-09-735-487-11	Sequence 11, Appl
23	1081.6	44.0	2601	4	US-09-735-487-13	Sequence 13, Appl
24	1081.6	44.0	4307	4	US-09-552-950-1	Sequence 1, Appli
25	1081.6	44.0	9719	4	US-09-700-304-1	Sequence 1, Appli
26	1080	44.0	9050	4	US-09-184-418C-7	Sequence 7, Appli
27	1075.2	43.8	7399	2	US-08-418-848A-9	Sequence 9, Appli
28	1075.2	43.8	9709	2	US-08-188-583-5	Sequence 5, Appli
29	1075.2	43.8	9709	3	US-08-388-353-1	Sequence 1, Appli
30	1075.2	43.8	9709	3	US-08-488-551B-1	Sequence 1, Appli
31	1075.2	43.8	9709	4	US-09-309-572-15	Sequence 15, Appl
32	1075.2	43.8	9709	4	US-09-718-096-15	Sequence 15, Appl
33	1075.2	43.8	12479	4	US-09-318-138-13	Sequence 13, Appl
34	1075.2	43.8	12494	3	US-08-935-312-13	Sequence 13, Appl
35	1075.2	43.8	12494	3	US-08-848-760B-33	Sequence 33, Appl
36	1075.2	43.8	15581	3	US-08-646-538-35	Sequence 35, Appl
37	1075.2	43.8	15581	3	US-09-503-222-35	Sequence 35, Appl
38	1074	43.7	3000	4	US-09-184-418C-74	Sequence 74, Appl
39	1070.6	43.6	8968	4	US-09-184-418C-1	Sequence 1, Appli
40	1065.6	43.4	4313	4	US-09-475-515-3	Sequence 3, Appli
41	1065.6	43.4	9737	2	US-08-944-449-7	Sequence 7, Appli
42	1065.6	43.4	9737	4	US-09-353-362-7	Sequence 7, Appli
43	1057.6	43.0	8954	4	US-09-184-418C-6	Sequence 6, Appli
44	1052.4	42.8	3017	4	US-09-184-418C-83	Sequence 83, Appl
45	1050.2	42.7	3011	4	US-09-184-418C-101	Sequence 101, App

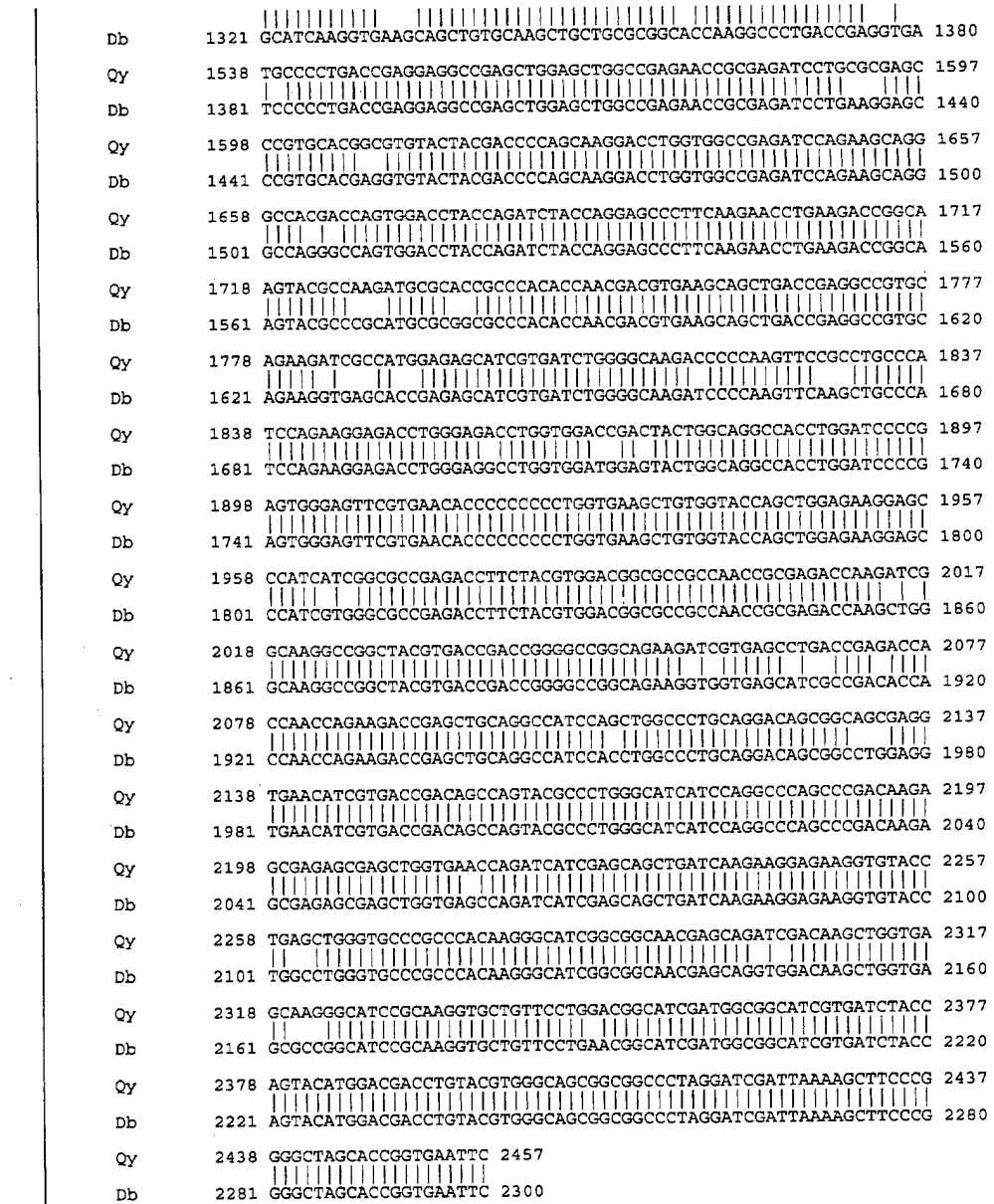
ALIGNMENTS

RESULT 1
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YMWM
US-09-475-515-83

Query Match 83.0%; Score 2040; DB 4; Length 2300;
Best Local Similarity 93.6%; Pred. No. 0;
Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              ||||||||||||||| |||||||||||||||||||| |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||| | ||||      ||||| || ||||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      ||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
```



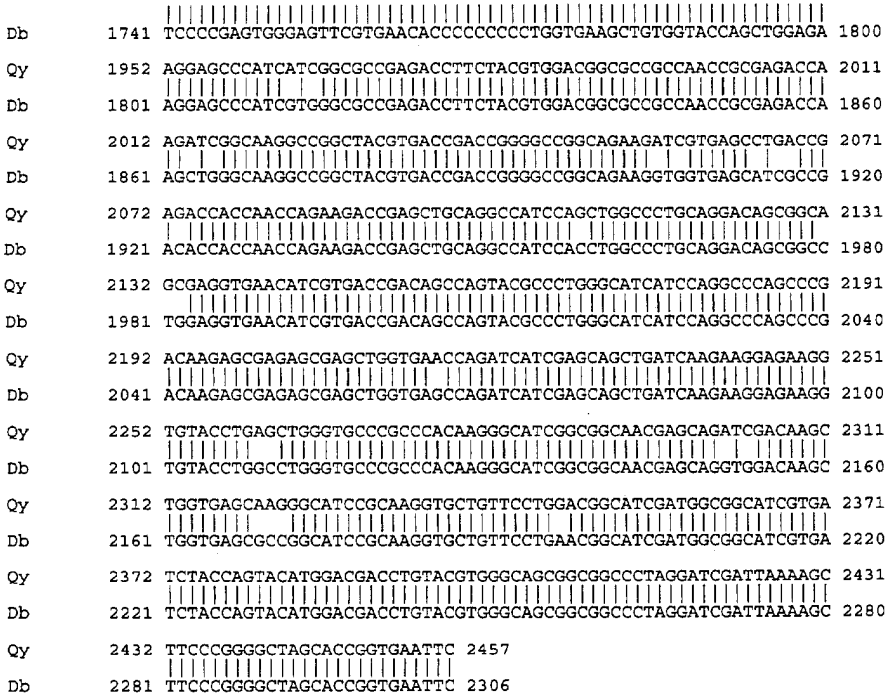


RESULT 2
US-09-475-515-82
; Sequence 82, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 2306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

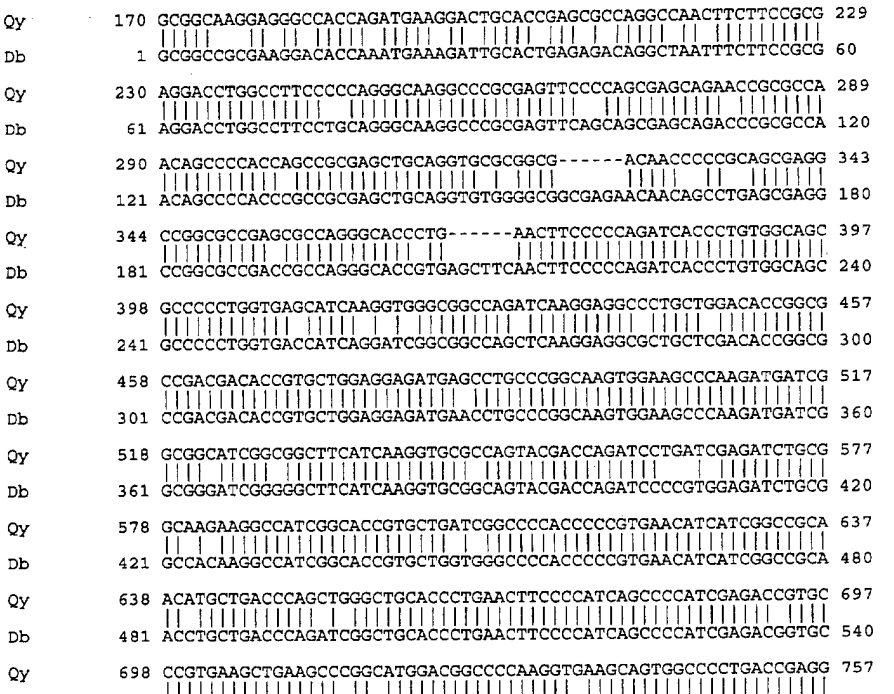
Query Match 82.2%; Score 2019.2; DB 4; Length 2306;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

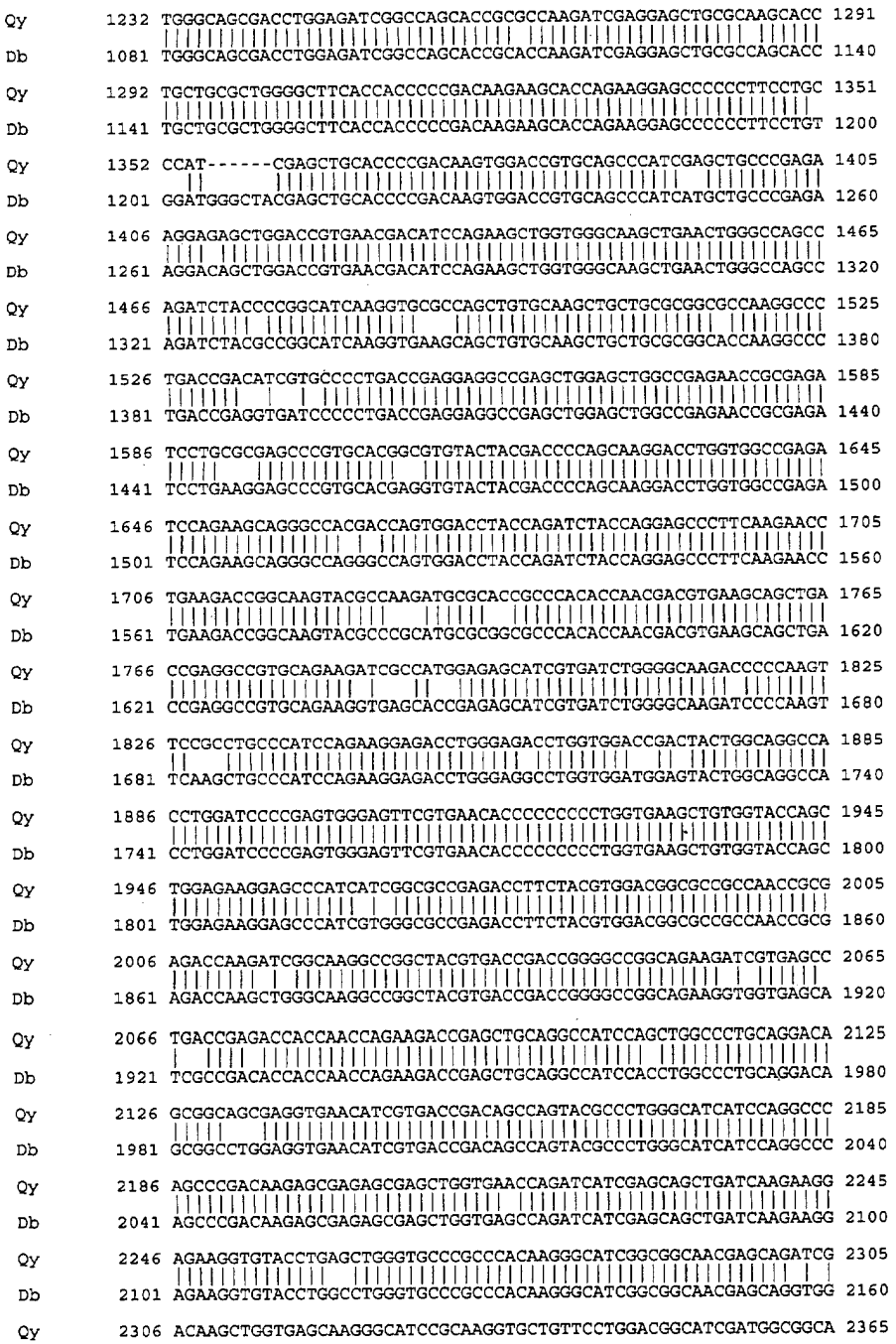
```
Qy   170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db     1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG  60

Qy   230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db    61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy   290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db   121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy   344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db   181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy   398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db   241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy   458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db   301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy   518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db   361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy   578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db   421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy   638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db   481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy   698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db   541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy   758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db   601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy   818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db   661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy   878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db   721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy   938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db   781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy   998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db   841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy  1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db   901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy  1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db   961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy  1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db  1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy  1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db  1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy  1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db  1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy  1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db  1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy  1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db  1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy  1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db  1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy  1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db  1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy  1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db  1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy  1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db  1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy  1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db  1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy  1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db  1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy  1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
Db  1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy  1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
```

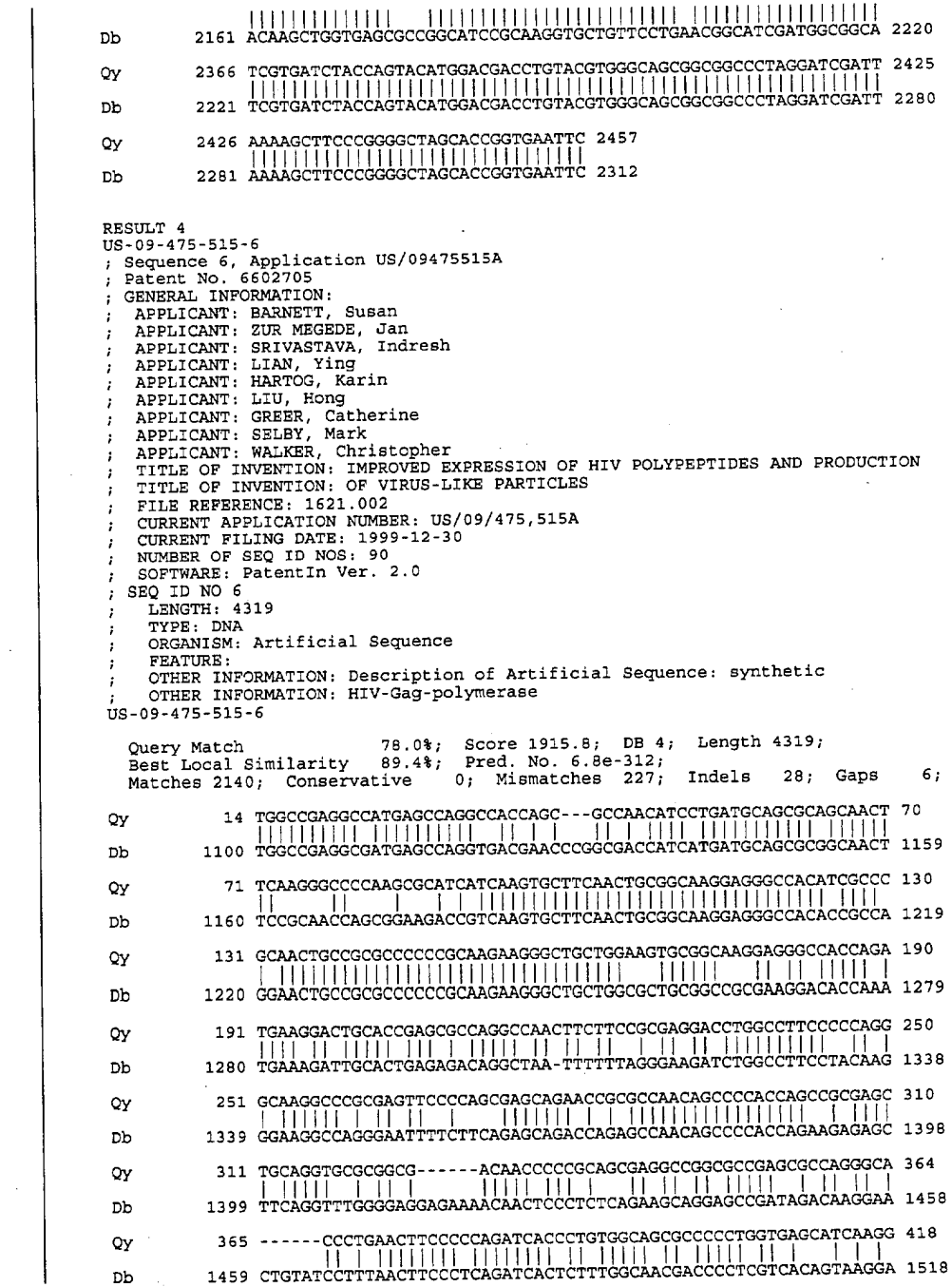


RESULT 3
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 84
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt(+)
US-09-475-515-84

Query Match 81.3%; Score 1998.4; DB 4; Length 2312;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

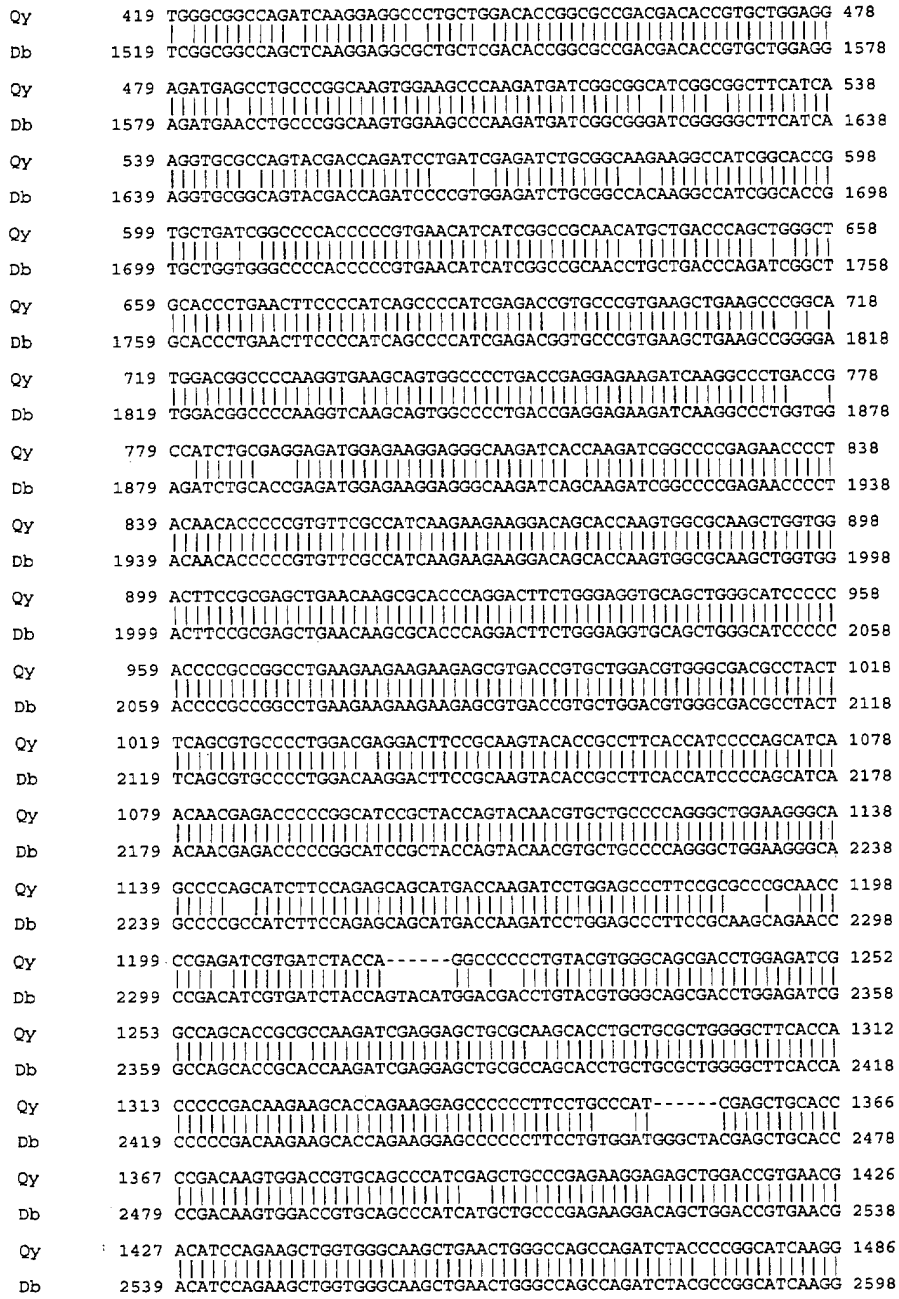


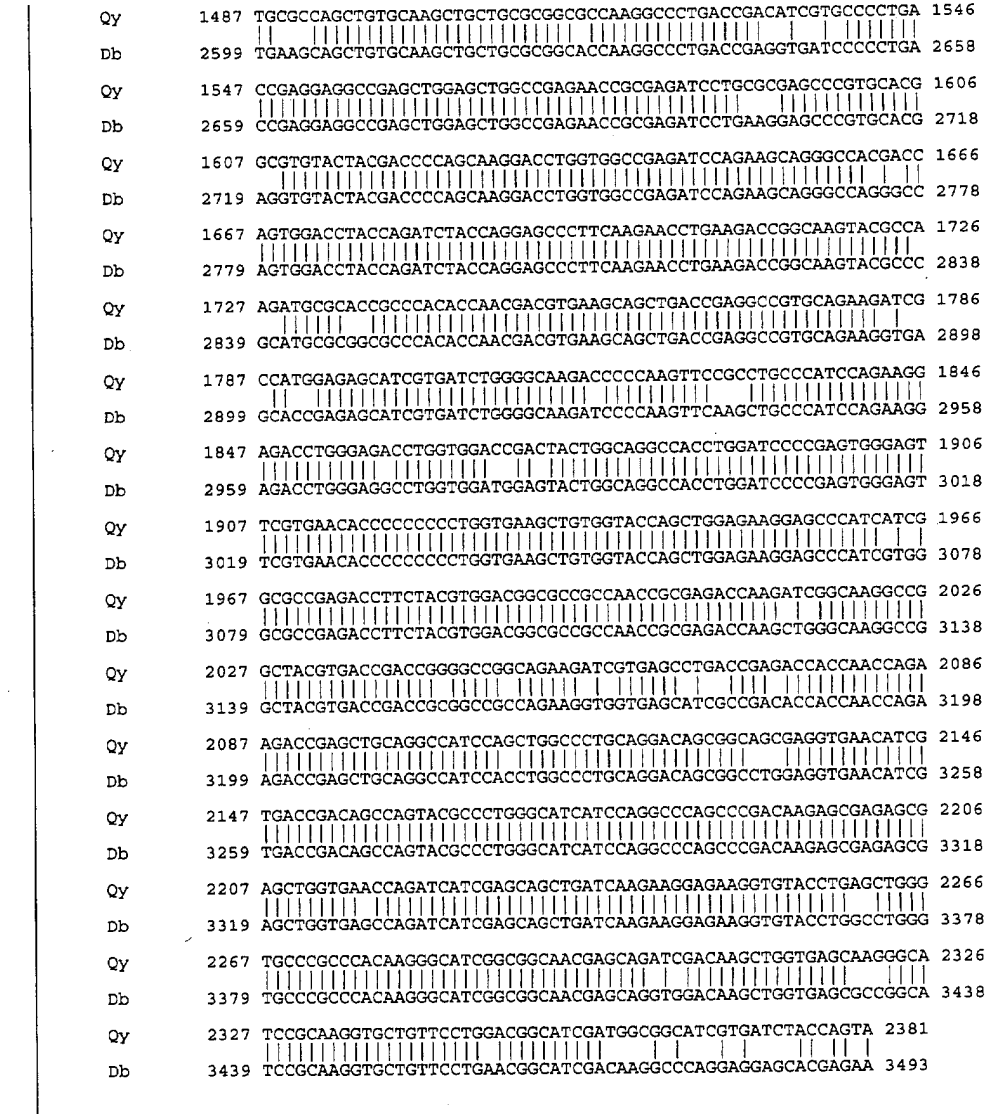




RESULT 4
US-09-475-515-6
; Sequence 6, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: HIV-Gag-polymerase
US-09-475-515-6

Query Match 78.0%; Score 1915.8; DB 4; Length 4319;
Best Local Similarity 89.4%; Pred. No. 6.8e-312;
Matches 2140; Conservative 0; Mismatches 227; Indels 28; Gaps 6;





RESULT 5
US-09-475-515-81
; Sequence 81, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine

```
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
;   LENGTH: 2299
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(+).proinact.RTopt.YMWM
US-09-475-515-81

  Query Match             76.2%;  Score 1872.8;  DB 4;  Length 2299;
  Best Local Similarity   89.3%;  Pred. No. 9.6e-305;
  Matches 2055;  Conservative    0;  Mismatches  232;  Indels   13;  Gaps    3;

Qy       170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db         1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG  59

Qy       230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db        60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy       290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db       120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy       344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db       180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy       398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db       240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy       458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db       300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy       518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db       360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy       578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db       420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy       638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db       480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy       698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db       540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy       758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db       600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy       818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db       660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy       878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db       720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy       938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db       780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy       998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db       840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy      1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db       900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy      1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db       960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy      1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCCTGTACGTGGGCA 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1079

Qy      1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1080 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1139

Qy      1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db       1140 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1199

Qy      1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1200 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1259

Qy      1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1260 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1319

Qy      1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1320 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1379

Qy      1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1380 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1439

Qy      1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1440 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1499

Qy      1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db       1500 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1559

Qy      1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db       1560 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1619

Qy      1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db       1620 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1679

Qy      1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db       1680 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1739

Qy      1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db       1740 AGTGGGAGTTCGTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1799

Qy      1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db       1800 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1859
```

```
Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db       1860 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1919

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db       1920 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1979

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db       1980 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2039

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db       2040 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2099

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db       2100 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2159

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2160 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2219

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2220 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2279

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2280 GGGCTAGCACCGGTGAATTC 2299
```

RESULT 6
US-09-475-515-80
; Sequence 80, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).proinact.RTopt.YM
US-09-475-515-80

Query Match 75.4%; Score 1852; DB 4; Length 2305;
Best Local Similarity 89.0%; Pred. No. 2.9e-301;
Matches 2052; Conservative 0; Mismatches 235; Indels 19; Gaps 4;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
```

```
Db         60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy        344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy       1178 AGCCCTTCCGCGCCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1079

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1080 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1139

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db       1140 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1199
```

```
Qy       1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db       1200 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1259

Qy       1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db       1260 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1319

Qy       1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db       1320 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1379

Qy       1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db       1380 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1439

Qy       1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db       1440 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1499

Qy       1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db       1500 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1559

Qy       1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db       1560 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1619

Qy       1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db       1620 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1679

Qy       1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
Db       1680 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1739

Qy       1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
Db       1740 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1799

Qy       1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011
Db       1800 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1859

Qy       2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071
Db       1860 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1919

Qy       2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131
Db       1920 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1979

Qy       2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191
Db       1980 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2039

Qy       2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251
Db       2040 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2099

Qy       2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311
Db       2100 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2159

Qy       2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371
Db       2160 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2219

Qy       2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431
Db       2220 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2279

Qy       2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457
Db       2280 TTCCCGGGGCTAGCACCGGTGAATTC 2305
```

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-SYNgp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 66.2%; Score 1626.6; DB 4; Length 4307;
Best Local Similarity 82.2%; Pred. No. 1.4e-263;
Matches 1951; Conservative 0; Mismatches 394; Indels 28; Gaps 6;

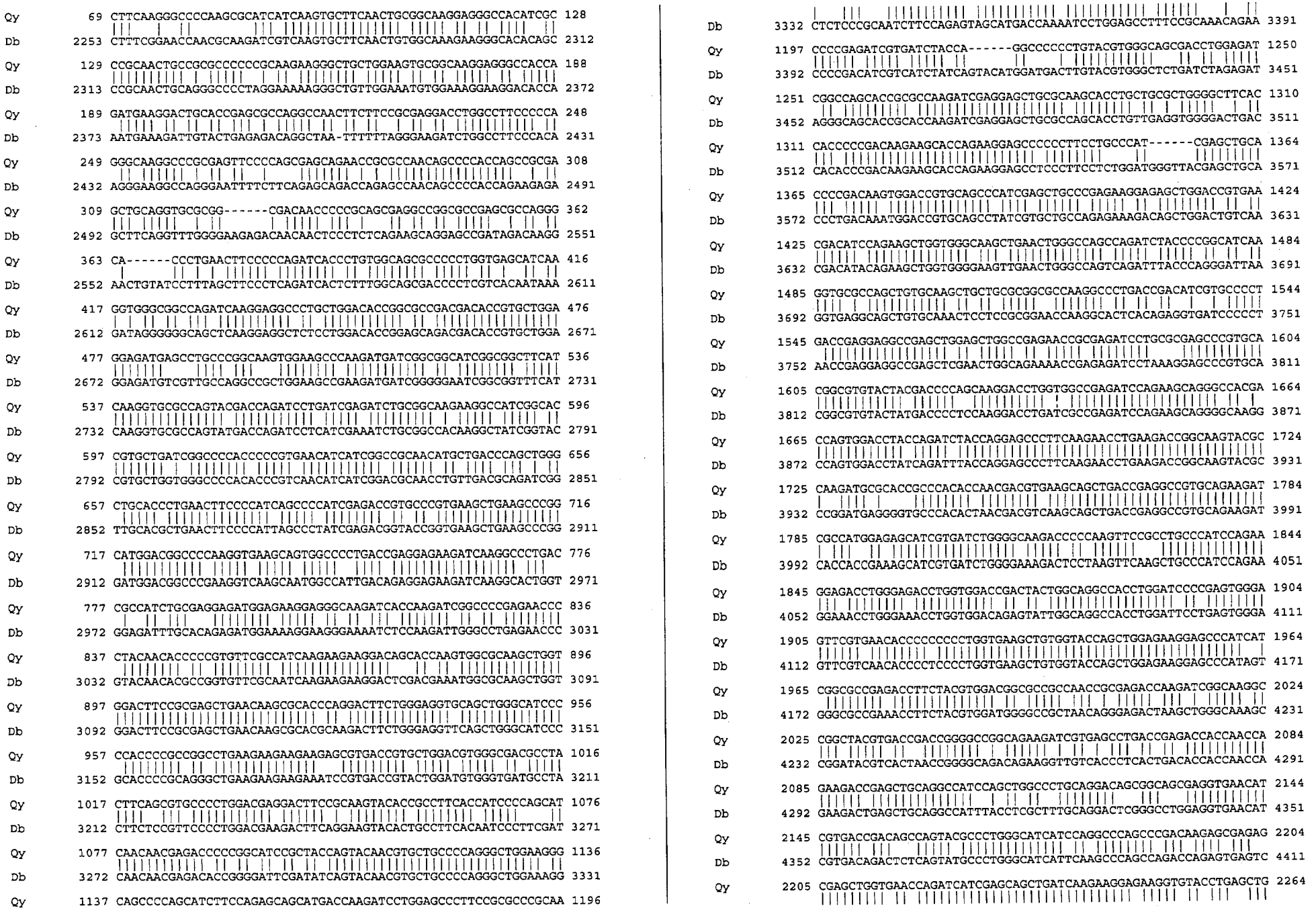
```
Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db       1086 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1145

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db       1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1205

Qy        129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db       1206 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCA 1265

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db       1266 GATGAAAGACTGTACGGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACA 1324

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db       1325 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCCGCGCCAACAGCCCCACCCGCCGCGA 1384

Qy        309 GCTGCAGGT------GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db       1385 GCTTCAGGTCTGGGGTCGCGACAACAACTCCCCCTCCGAAGCAGGAGCCGACCGCCAGGG 1444

Qy        363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db       1445 CACGGTGTCCTTCAACTTCCCTCAGGTCACGCTTTGGCAGCGACCCCTCGTCACCATCAA 1504

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db       1505 GATCGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1564

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db       1565 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1624

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCAAGAAGGCCATCGGCAC 596
Db       1625 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1684

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db       1685 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1744
```

```
Qy       657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db      1745 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1804

Qy       717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db      1805 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1864

Qy       777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db      1865 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 1924

Qy       837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db      1925 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 1984

Qy       897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db      1985 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2044

Qy       957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db      2045 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2104

Qy      1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db      2105 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2164

Qy      1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db      2165 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2224

Qy      1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db      2225 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2284

Qy      1197 CCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db      2285 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2344

Qy      1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
Db      2345 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2404

Qy      1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCA 1364
Db      2405 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2464

Qy      1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424
Db      2465 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2524

Qy      1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484
Db      2525 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2584

Qy      1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544
Db      2585 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2644

Qy      1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604
Db      2645 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2704

Qy      1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664
Db      2705 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2764

Qy      1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
Db      2765 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2824

Qy      1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
Db      2825 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2884

Qy      1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
Db      2885 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2944

Qy      1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
Db      2945 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3004

Qy      1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
Db      3005 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3064

Qy      1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db      3065 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3124

Qy      2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
Db      3125 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3184

Qy      2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
Db      3185 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3244

Qy      2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
Db      3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3304

Qy      2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
Db      3305 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3364

Qy      2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
Db      3365 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3424

Qy      2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357
Db      3425 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3457
```

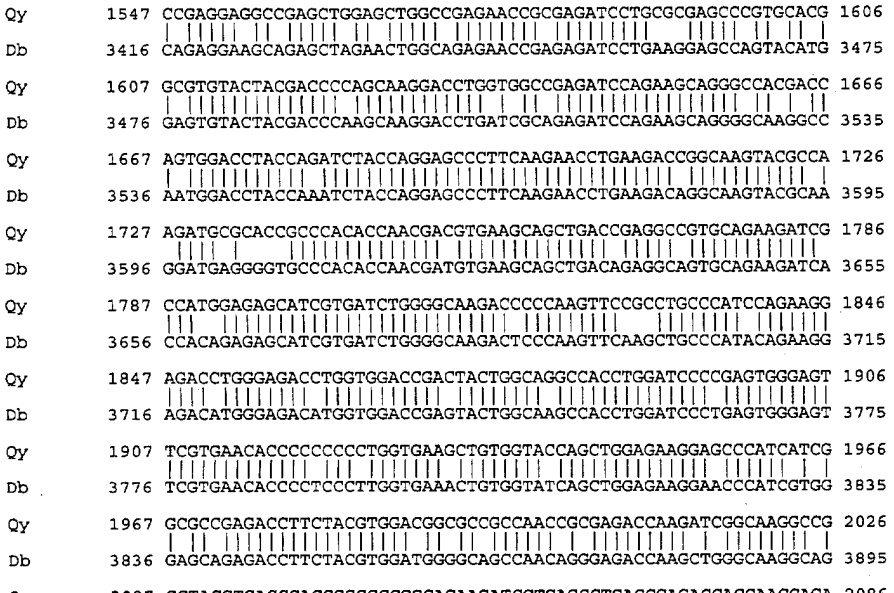
RESULT 8
US-09-552-950-5
; Sequence 5, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 9772
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:pSYNGP
US-09-552-950-5

Query Match 65.1%; Score 1599.4; DB 4; Length 9772;
Best Local Similarity 81.5%; Pred. No. 5.5e-259;
Matches 1934; Conservative 0; Mismatches 411; Indels 28; Gaps 6;

```
Qy        12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db      2193 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 2252
```



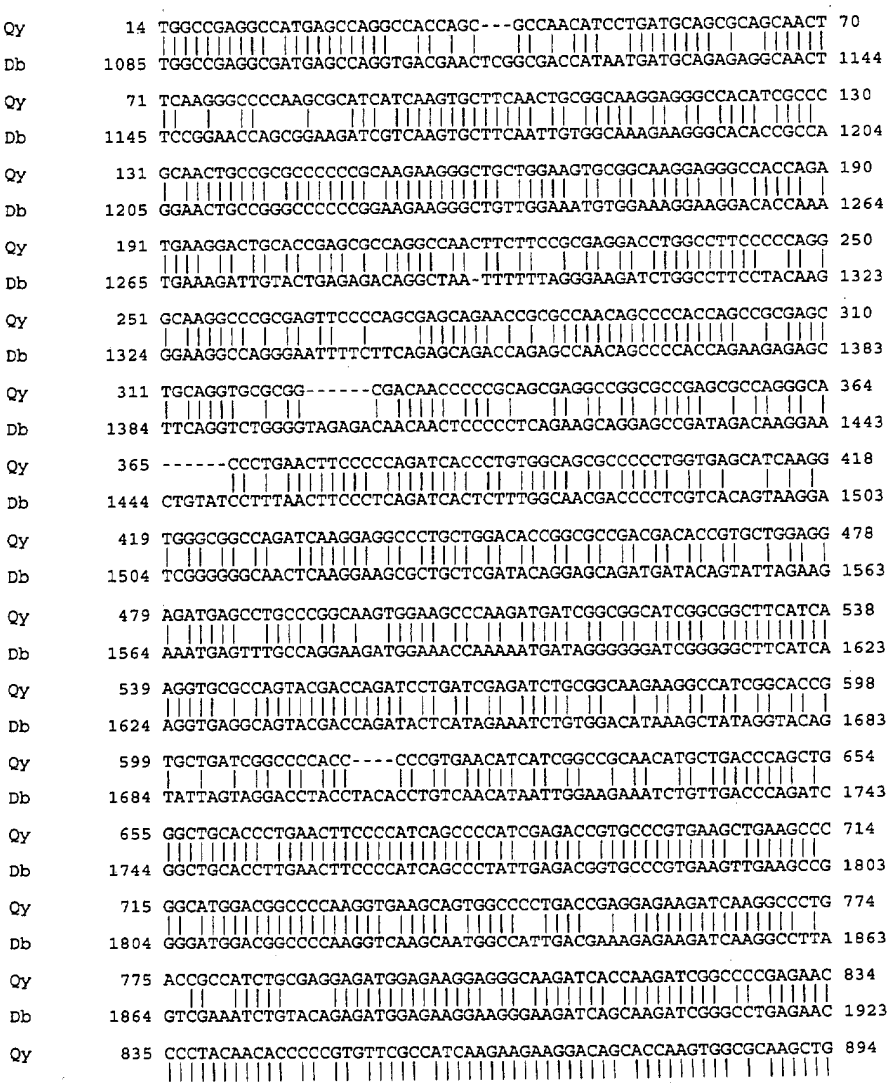
```
Db       4412 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 4471

Qy       2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324

Db       4472 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 4531

Qy       2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357

Db       4532 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 4564


RESULT 9
US-09-872-733A-6
; Sequence 6, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733A
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 8366
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  DNA sequence
;   OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
;   OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
;   OTHER INFORMATION: resistance gene
US-09-872-733A-6

  Query Match             62.6%;  Score 1539.2;  DB 4;  Length 8366;
  Best Local Similarity   79.9%;  Pred. No. 6.2e-249;
  Matches 1896;  Conservative    0;  Mismatches  448;  Indels   28;  Gaps    6;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70

Db       1857 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1916

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130

Db       1917 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1976

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190

Db       1977 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 2036

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250

Db       2037 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 2095

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310

Db       2096 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 2155

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364

Db       2156 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 2215

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418

Db       2216 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 2275

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478

Db       2276 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 2335

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538

Db       2336 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 2395

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598

Db       2396 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 2455

Qy        599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658

Db       2456 TATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCT 2515

Qy        659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718

Db       2516 GCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGA 2575

Qy        719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778

Db       2576 TGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCG 2635

Qy        779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838

Db       2636 AAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCT 2695

Qy        839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898

Db       2696 ACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGG 2755

Qy        899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958

Db       2756 ACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCAC 2815

Qy        959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018

Db       2816 ATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACT 2875

Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078

Db       2876 TCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCA 2935

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138

Db       2936 ACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGAT 2995

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198

Db       2996 CACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACC 3055

Qy       1199 CCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCG 1252

Db       3056 CAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCG 3115

Qy       1253 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1312

Db       3116 GGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCA 3175

Qy       1313 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCCTGCCCAT------CGAGCTGCACC 1366

Db       3176 CACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATC 3235

Qy       1367 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1426

Db       3236 CTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACG 3295

Qy       1427 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1486

Db       3296 ACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAG 3355

Qy       1487 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1546

Db       3356 TTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGA 3415
```



RESULT 10
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

Query Match 61.2%; Score 1503.8; DB 4; Length 4338;
Best Local Similarity 79.8%; Pred. No. 4.7e-243;
Matches 1896; Conservative 0; Mismatches 447; Indels 34; Gaps 9;



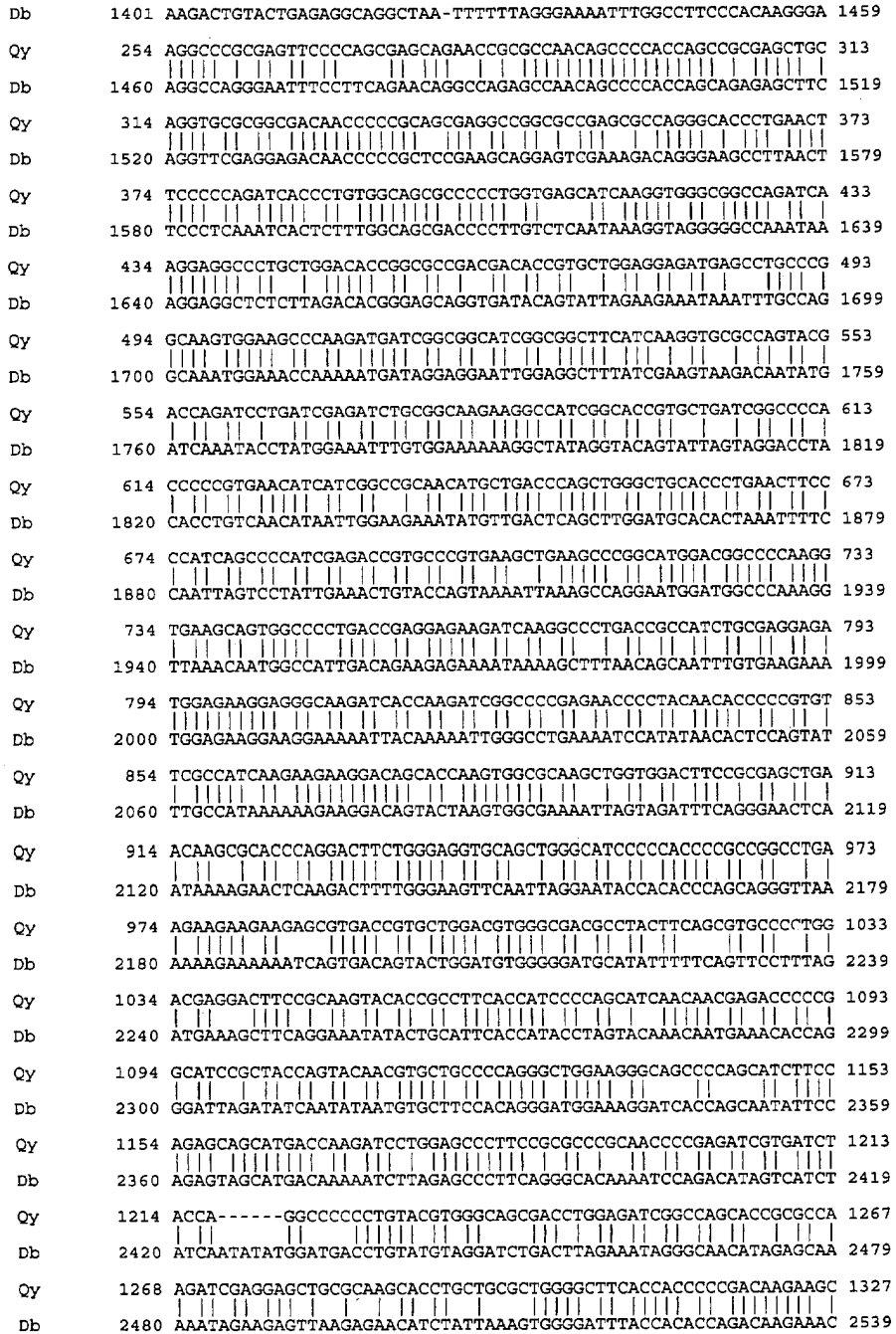
```
Db       1924 CCCTACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTG 1983
Qy        895 GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954
Db       1984 GTGGACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATC 2043
Qy        955 CCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC 1014
Db       2044 CCACATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCC 2103
Qy       1015 TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGC 1074
Db       2104 TACTTCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGC 2163
Qy       1075 ATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAG 1134
Db       2164 ATCAACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAG 2223
Qy       1135 GGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGC 1194
Db       2224 GGATCACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAA 2283
Qy       1195 AACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAG 1248
Db       2284 AACCCAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAG 2343
Qy       1249 ATC-GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTT 1307
Db       2344 ATCGGGGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACT 2403
Qy       1308 CACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCT 1361
Db       2404 GACCACACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACT 2463
Qy       1362 GCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGT 1421
Db       2464 GCATCCTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGT 2523
Qy       1422 GAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCAT 1481
Db       2524 GAACGACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCAT 2583
Qy       1482 CAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCC 1541
Db       2584 CAAAGTTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCC 2643
Qy       1542 CCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGT 1601
Db       2644 ACTGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGT 2703
Qy       1602 GCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCA 1661
Db       2704 ACATGGAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCA 2763
Qy       1662 CGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTA 1721
Db       2764 AGGCCAATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTA 2823
Qy       1722 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1781
Db       2824 CGCAAGGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAA 2883
Qy       1782 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1841
Db       2884 GATCACCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943
Qy       1842 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1901
Db       2944 GAAGGAGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTG 3003
Qy       1902 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1961
Db       3004 GGAGTTCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCAT 3063
```

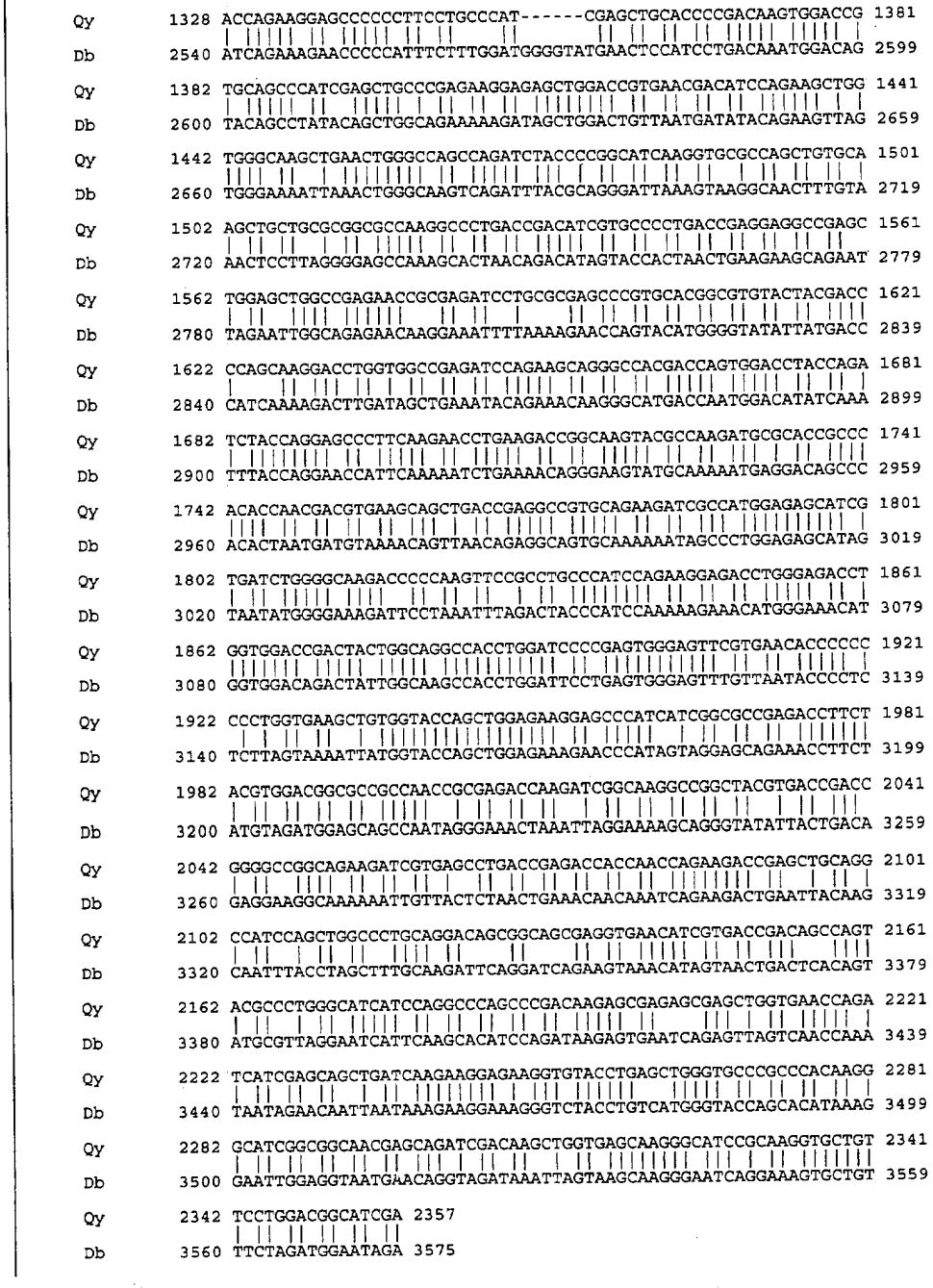
```
Qy       1962 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2021
Db       3064 CGTGGGAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAA 3123
Qy       2022 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2081
Db       3124 GGCAGGCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAA 3183
Qy       2082 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2141
Db       3184 CCAGAAGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAA 3243
Qy       2142 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2201
Db       3244 CATCGTGACAGACTCACAGTACG-CATGGGCATCATCCAAGCACAACCAGACCAATCCGA 3302
Qy       2202 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2261
Db       3303 GTCAGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGC 3362
Qy       2262 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2321
Db       3363 ATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGC 3422
Qy       2322 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db       3423 TGGGATCCGGAAGGTGCTGTTCCTGGACGGGATCGAT 3459
```

RESULT 11
US-09-184-418C-8
; Sequence 8, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Peng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 8
; LENGTH: 9010
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=96ZM651; 137..1621:"gag"; 1426..4425:"pol";
; OTHER INFORMATION: 4370..4948:"vif"; 4888..5178:"vpr";
; OTHER INFORMATION: 5159..5373-7734..7824:"tat"; 5298..5373-7734..7981:"rev";
; OTHER INFORMATION: 5387..5647:"vpu"; 5565..8171:"env"; 8173..8793:"nef"
US-09-184-418C-8

Query Match 48.0%; Score 1178.4; DB 4; Length 9010;
Best Local Similarity 69.7%; Pred. No. 1.2e-188;
Matches 1642; Conservative 0; Mismatches 701; Indels 13; Gaps 3;

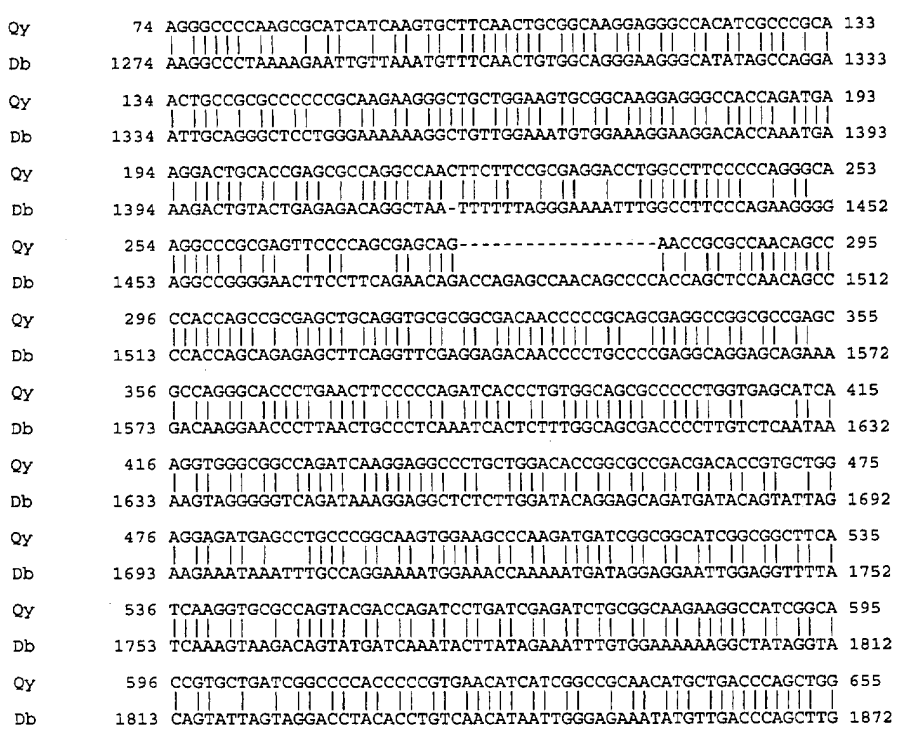
```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1221 TGGCTGAGGCAATGAGCCAAACAAATAGTGTAAACATACTGATGCAGAAAAGCAATTTTA 1280
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1281 AAGGAAATAAAAGAATGGTTAAATGTTTTAACTGTGGTAAGGAAGGGCACATAGCCAGAA 1340
Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1341 ATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAGGGACACCAAATGA 1400
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
```

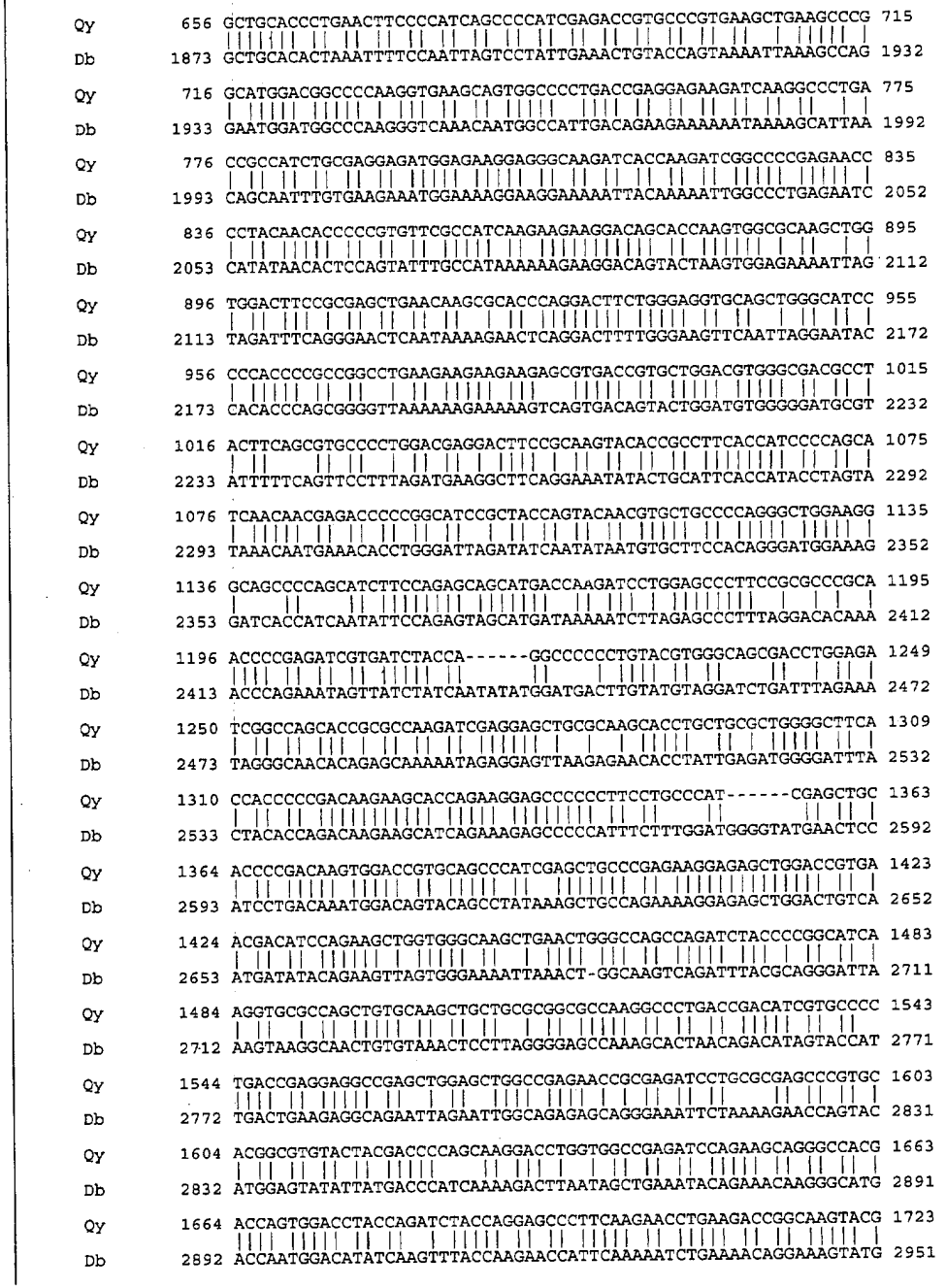

Db 1401 AAGACTGTACTGAGAGGCAGGCTAA-TTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGA 1459
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 1460 AGGCCAGGGAATTTCCTTCAGAACAGGCCAGAGCCAACAGCCCCACCAGCAGAGAGCTTC 1519
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1520 AGGTTCGAGGAGACAACCCCCGCTCCGAAGCAGGAGTCGAAAGACAGGGAAGCCTTAACT 1579
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1580 TCCCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAGGTAGGGGGCCAAATAA 1639
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1640 AGGAGGCTCTCTTAGACACGGGAGCAGGTGATACAGTATTAGAAGAAATAAATTTGCCAG 1699
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1700 GCAAATGGAAACCAAAAATGATAGGAGGAATTGGAGGCTTTATCGAAGTAAGACAATATG 1759
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 1760 ATCAAATACCTATGGAAATTTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGACCTA 1819
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 1820 CACCTGTCAACATAATTGGAAGAAATATGTTGACTCAGCTTGGATGCACACTAAATTTTC 1879
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 1880 CAATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAGG 1939
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 1940 TTAAACAATGGCCATTGACAGAAGAGAAAATAAAAGCTTTAACAGCAATTTGTGAAGAAA 1999
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2000 TGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGAAAATCCATATAACACTCCAGTAT 2059
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2060 TTGCCATAAAAAAGAAGGACAGTACTAAGTGGCGAAAATTAGTAGATTTCAGGGAACTCA 2119
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2120 ATAAAAGAACTCAAGACTTTTGGGAAGTTCAATTAGGAATACCACACCCAGCAGGGTTAA 2179
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2180 AAAAGAAAAAATCAGTGACAGTACTGGATGTGGGGGATGCATATTTTTCAGTTCCTTTAG 2239
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2240 ATGAAAGCTTCAGGAAATATACTGCATTCACCATACCTAGTACAAACAATGAAACACCAG 2299
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2300 GGATTAGATATCAATATAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCC 2359
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2360 AGAGTAGCATGACAAAAATCTTAGAGCCCTTCAGGGCACAAAATCCAGACATAGTCATCT 2419
Qy 1214 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1267
Db 2420 ATCAATATATGGATGACCTGTATGTAGGATCTGACTTAGAAATAGGGCAACATAGAGCAA 2479
Qy 1268 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1327
Db 2480 AAATAGAAGAGTTAAGAGAACATCTATTAAAGTGGGGATTTACCACACCAGACAAGAAAC 2539

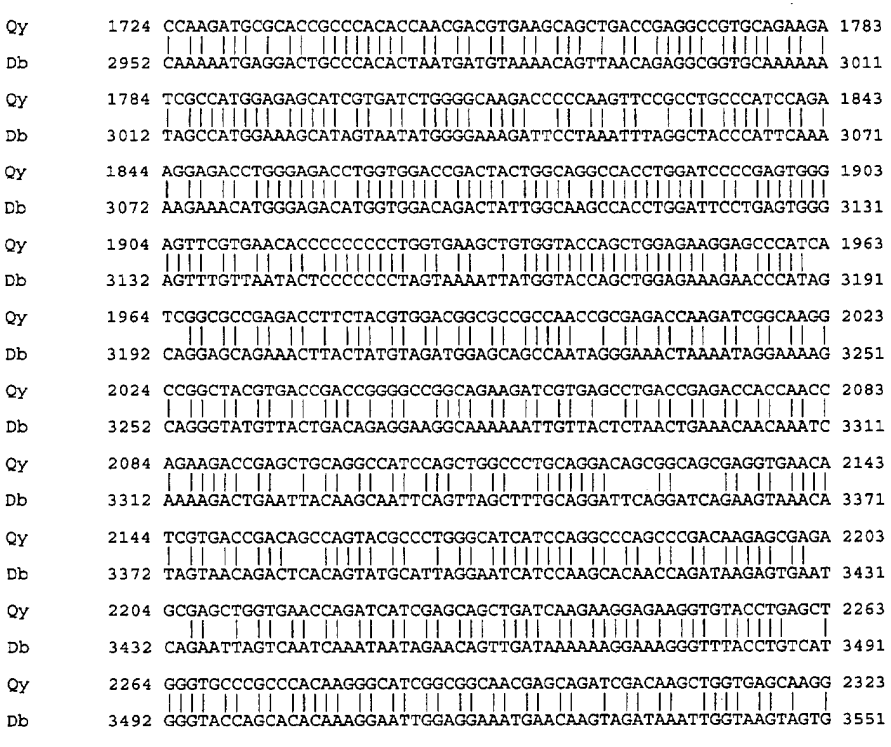

Qy 1328 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 1381
Db 2540 ATCAGAAAGAACCCCCATTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACAG 2599
Qy 1382 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1441
Db 2600 TACAGCCTATACAGCTGGCAGAAAAAGATAGCTGGACTGTTAATGATATACAGAAGTTAG 2659
Qy 1442 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1501
Db 2660 TGGGAAAATTAAACTGGGCAAGTCAGATTTACGCAGGGATTAAAGTAAGGCAACTTTGTA 2719
Qy 1502 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1561
Db 2720 AACTCCTTAGGGGAGCCAAAGCACTAACAGACATAGTACCACTAACTGAAGAAGCAGAAT 2779
Qy 1562 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1621
Db 2780 TAGAATTGGCAGAGAACAAGGAAATTTTAAAAGAACCAGTACATGGGGTATATTATGACC 2839
Qy 1622 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1681
Db 2840 CATCAAAAGACTTGATAGCTGAAATACAGAAACAAGGGCATGACCAATGGACATATCAAA 2899
Qy 1682 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1741
Db 2900 TTTACCAGGAACCATTCAAAAATCTGAAAACAGGGAAGTATGCAAAAATGAGGACAGCCC 2959
Qy 1742 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1801
Db 2960 ACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCCCTGGAGAGCATAG 3019
Qy 1802 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1861
Db 3020 TAATATGGGGAAAGATTCCTAAATTTAGACTACCCATCCAAAAAGAAACATGGGAAACAT 3079
Qy 1862 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1921
Db 3080 GGTGGACAGACTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACCCCTC 3139
Qy 1922 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1981
Db 3140 TCTTAGTAAAATTATGGTACCAGCTGGAGAAAGAACCCATAGTAGGAGCAGAAACCTTCT 3199
Qy 1982 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2041
Db 3200 ATGTAGATGGAGCAGCCAATAGGGAAACTAAATTAGGAAAAGCAGGGTATATTACTGACA 3259
Qy 2042 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2101
Db 3260 GAGGAAGGCAAAAAATTGTTACTCTAACTGAAACAACAAATCAGAAGACTGAATTACAAG 3319
Qy 2102 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2161
Db 3320 CAATTTACCTAGCTTTGCAAGATTCAGGATCAGAAGTAAACATAGTAACTGACTCACAGT 3379
Qy 2162 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2221
Db 3380 ATGCGTTAGGAATCATTCAAGCACATCCAGATAAGAGTGAATCAGAGTTAGTCAACCAAA 3439
Qy 2222 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2281
Db 3440 TAATAGAACAATTAATAAAGAAGGAAAGGGTCTACCTGTCATGGGTACCAGCACATAAAG 3499
Qy 2282 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2341
Db 3500 GAATTGGAGGTAATGAACAGGTAGATAAATTAGTAAGCAAGGGAATCAGGAAAGTGCTGT 3559
Qy 2342 TCCTGGACGGCATCGA 2357
Db 3560 TTCTAGATGGAATAGA 3575

RESULT 12
US-09-184-418C-9
; Sequence 9, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 9
; LENGTH: 8972
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=96ZM751.3; 137.1632:gag; 1419.4435:pol;
; OTHER INFORMATION: 4380.4958:vif; 4898.5188:vpr; 5169.7814:tat;
; OTHER INFORMATION: 5308.7938:rev; 5407.5667:vpu; 5585.8128:env;
; OTHER INFORMATION: 8130.8753:nef
US-09-184-418C-9

Query Match 46.7%; Score 1147.4; DB 4; Length 8972;
Best Local Similarity 69.3%; Pred. No. 1.8e-183;
Matches 1647; Conservative 0; Mismatches 696; Indels 32; Gaps 5;

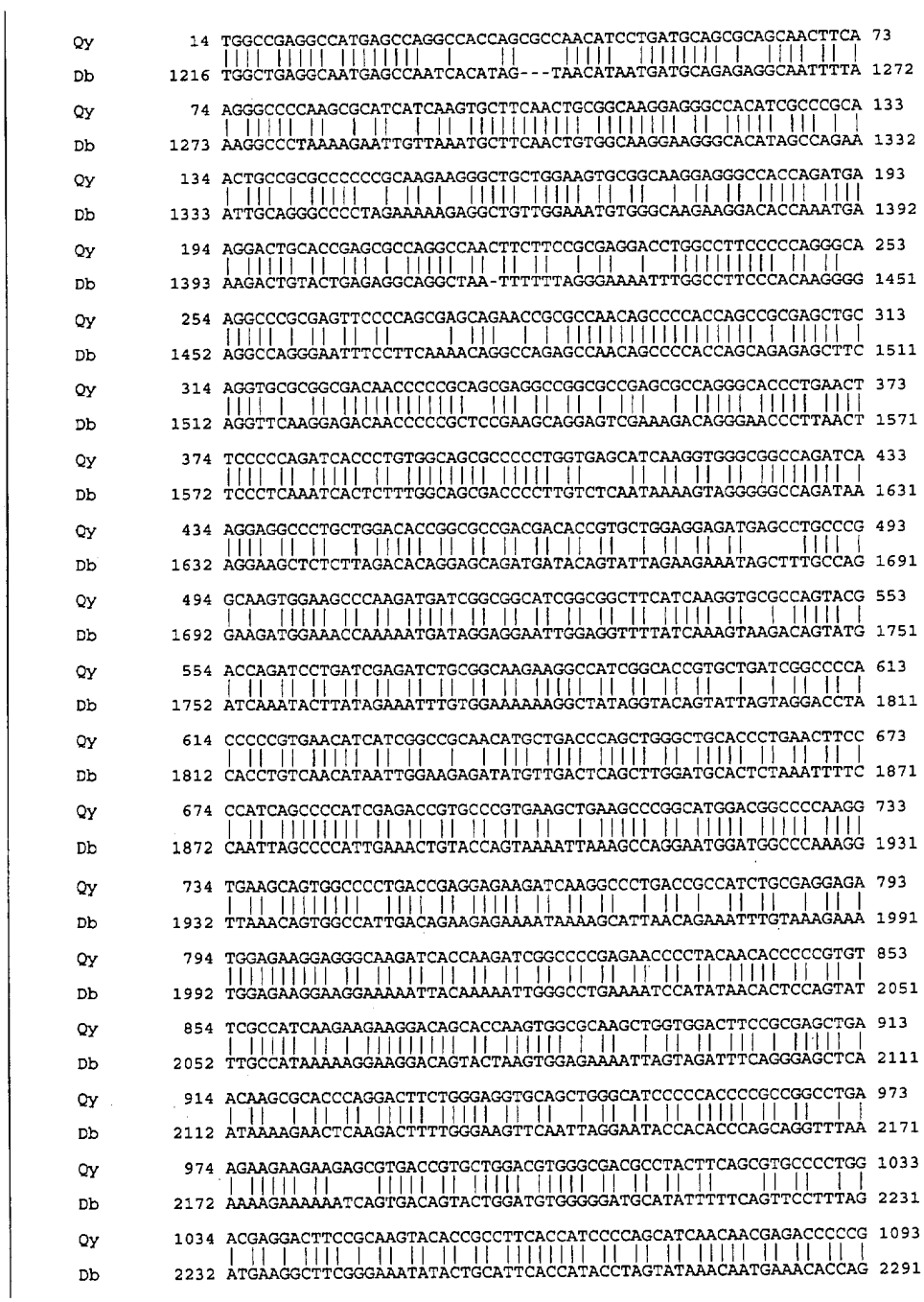






RESULT 13
US-09-184-418C-11
; Sequence 11, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 11
; LENGTH: 8959
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=94IN476.104; 138.1613:"gag";
; OTHER INFORMATION: 1418.4428:"pol"; 4361.4939:"vif"; 4879.5169:"vpr";
; OTHER INFORMATION: 5150.7782:tat"; 5289.7939:"rev"; 5378.5638:"vpu";
; OTHER INFORMATION: 5556.8129:"env"; 8131.8754:nef
US-09-184-418C-11

Query Match 46.4%; Score 1140.2; DB 4; Length 8959;
Best Local Similarity 69.2%; Pred. No. 2.9e-182;
Matches 1632; Conservative 0; Mismatches 708; Indels 17; Gaps 5;



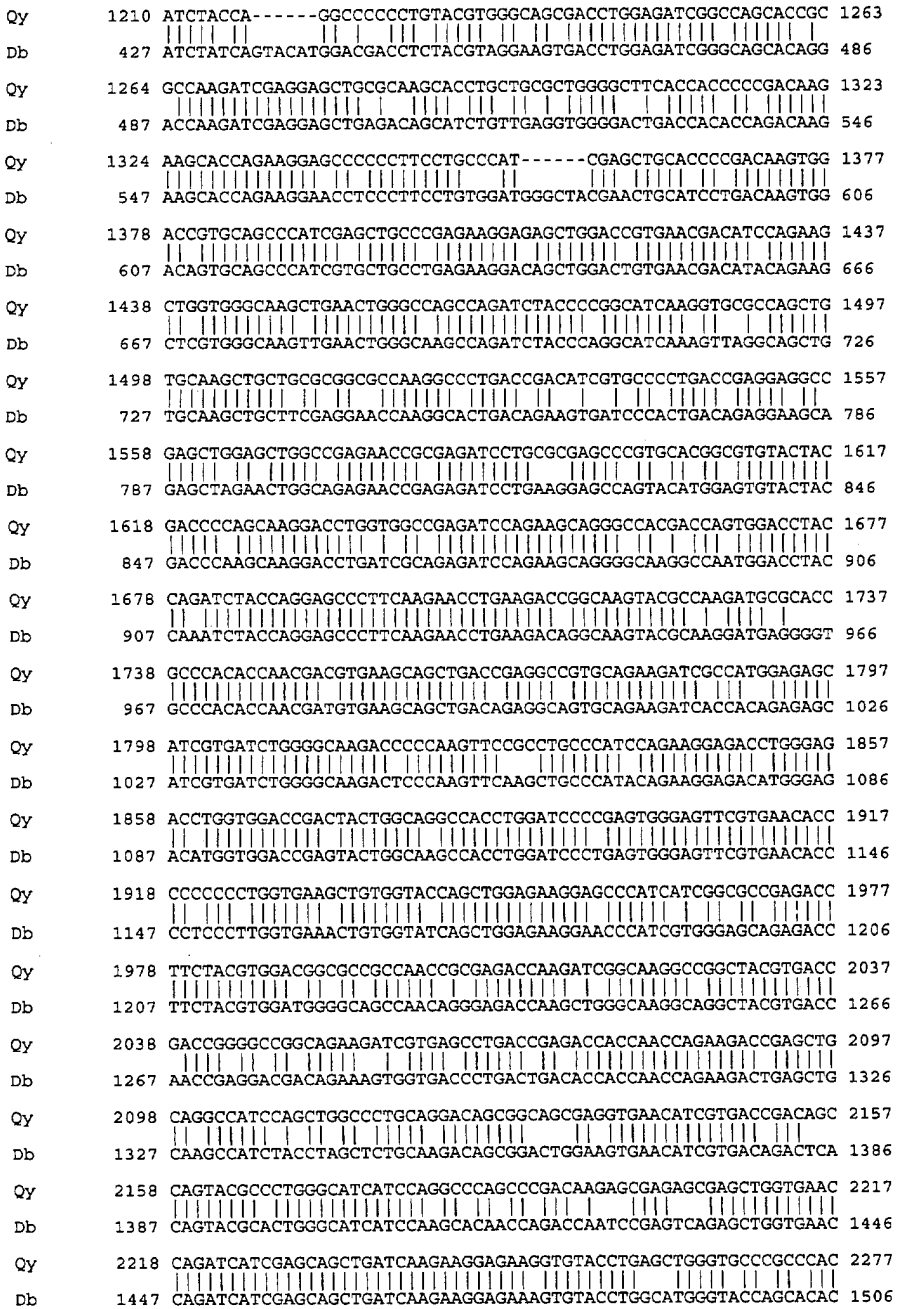
```
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db       2292 GGATTAGATATCAATATAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCC 2351

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db       2352 AGAGTAGCATGACAAAAATCTTAGAGCCCTTTAGGGCACGAAATCCAAAAATAGTCATCT 2411

Qy       1214 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1267
Db       2412 ATCAATATATGGATGACTTGTATGTAGGGTCTGACTTAGAAATAGGGCATCATAGAGCAA 2471

Qy       1268 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1327
Db       2472 AAATAGAGGAGTTAAGAGCACATCTATTAAAGTGGGGATTCACCACACCAGATAAGAAAC 2531

Qy       1328 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 1381
Db       2532 ATCAGAAAGAACCCCCATTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACAG 2591

Qy       1382 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1441
Db       2592 TACAGCCTATAAAGCTGCCAGAAAAGGATAGCTGGACTGTCAATGATATACAGAAGTTAG 2651

Qy       1442 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1501
Db       2652 TGGGAAAATTAAACTGGGCAAGTCAGATTTACCCAGGGATTAAAGTGAGGCAACTTTGTA 2711

Qy       1502 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1561
Db       2712 AACTCCTTAGGGGGGCCAAAGCACTAACAGACATAGTACCACTAACTGAAGAAGCAGAAT 2771

Qy       1562 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1621
Db       2772 TAGAATTAGCAGAGAACAGGGAAATTCTAAAAGAGCCAGTACATGGAGTATATTATGACC 2831

Qy       1622 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1681
Db       2832 CATCAAAAGACTTAATAGCTGAAATACAGAAACAGGGGCATGACCAATGGACATATCAAA 2891

Qy       1682 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1741
Db       2892 TTTACCAAGAACCATTCAAAAATCTGAAAACAGGGAAGTATGCAAAAATGAGGACTGCTC 2951

Qy       1742 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1801
Db       2952 ACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCCATAGAAAGCATAG 3011

Qy       1802 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1861
Db       3012 TAATAT-GGGAAAGACCCCTAAATTTAGACTACCCATCCAAAAAGAAACGTGGGAGACAT 3070

Qy       1862 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1921
Db       3071 GGTGGACAGACTATTGGCAGGCCACCTGGATTCCTGATTGGGAGTTTGTTAATACCCCTC 3130

Qy       1922 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1981
Db       3131 CCCTAGTAAAATTATGGTACCAGCTAGAAAAAGAACCCATAGTAGGAGCAGAAACTTTCT 3190

Qy       1982 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2041
Db       3191 ATGTAGATGGAGCAGCTAATAGGGAAACTAAAGTAGGAAAAGCAGGGTATGTTACTGACA 3250

Qy       2042 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2101
Db       3251 GAGGAAGGCAGAAAATTGTTTCTTTAACTGAAACAACAAATCAGAAGACTGAATTGCAAG 3310

Qy       2102 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2161
Db       3311 CAATTCAGCTAGCTTTGCAAGATTCAGGAACAGAAGTAAACATAGTAACAGACTCACAGT 3370

Qy       2162 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2221
```

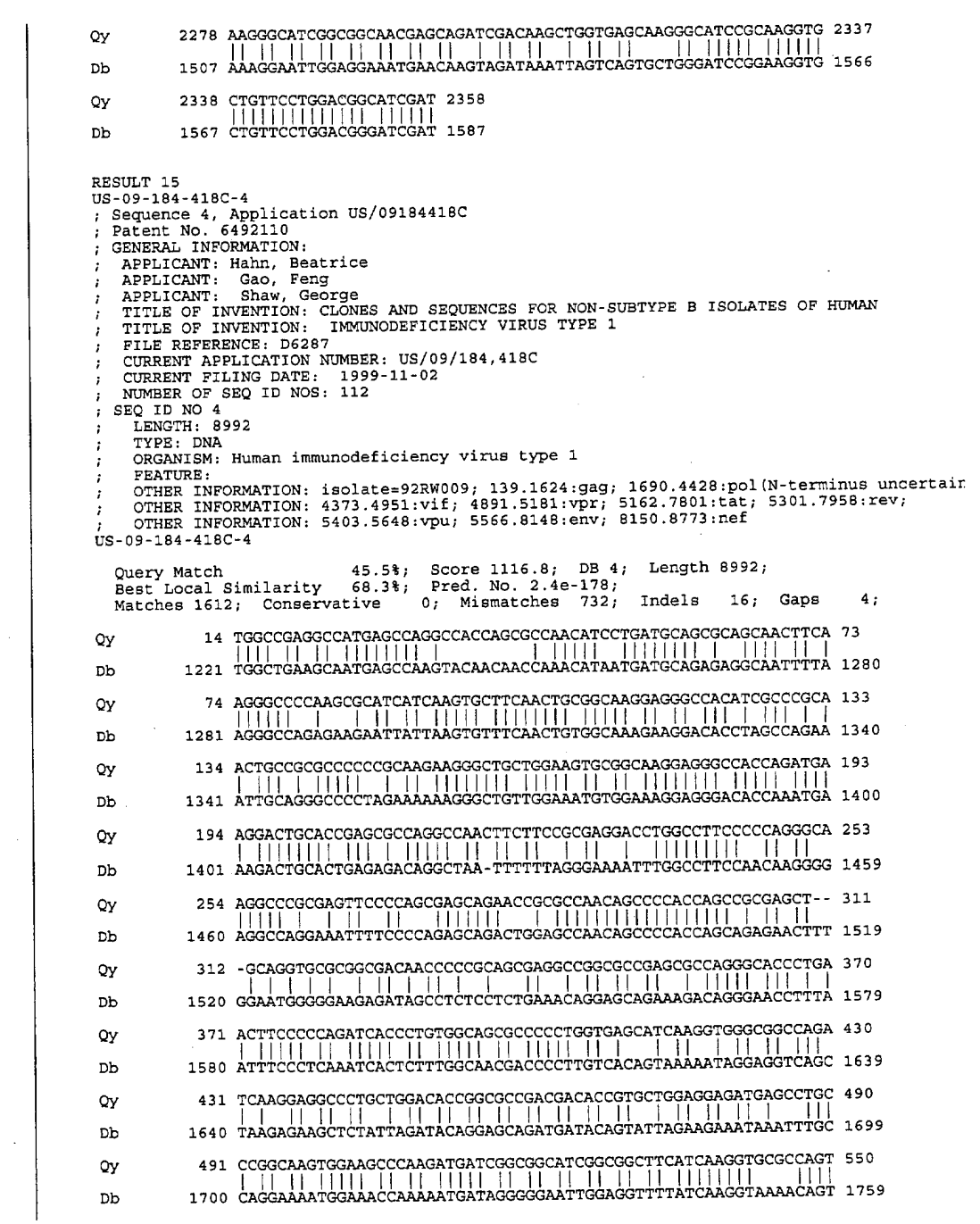
```
Db       3371 ATGCATTAGGAATCATTCAAGCACAACCAGATAAAAGTGAATCAGAGTTAGTCAACCAAA 3430

Qy       2222 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2281
Db       3431 TAATAGAACAATTAATAAACAAAGAAAGAGTCTATCTGTCATGGGTACCAGCACATAAAG 3490

Qy       2282 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2341
Db       3491 GAATTGGAGGGAATGAACAAGTAGATAGATTAGTAAGTAGTGGAATTAGGAAAGTACTGT 3550

Qy       2342 TCCTGGACGGCATCGAT 2358
Db       3551 TTCTAGATGGGATAGAT 3567
```

RESULT 14
US-09-872-733A-3
; Sequence 3, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Pol gene
US-09-872-733A-3

Query Match 46.3%; Score 1137; DB 4; Length 2467;
Best Local Similarity 83.4%; Pred. No. 8.6e-182;
Matches 1319; Conservative 0; Mismatches 250; Indels 12; Gaps 2;

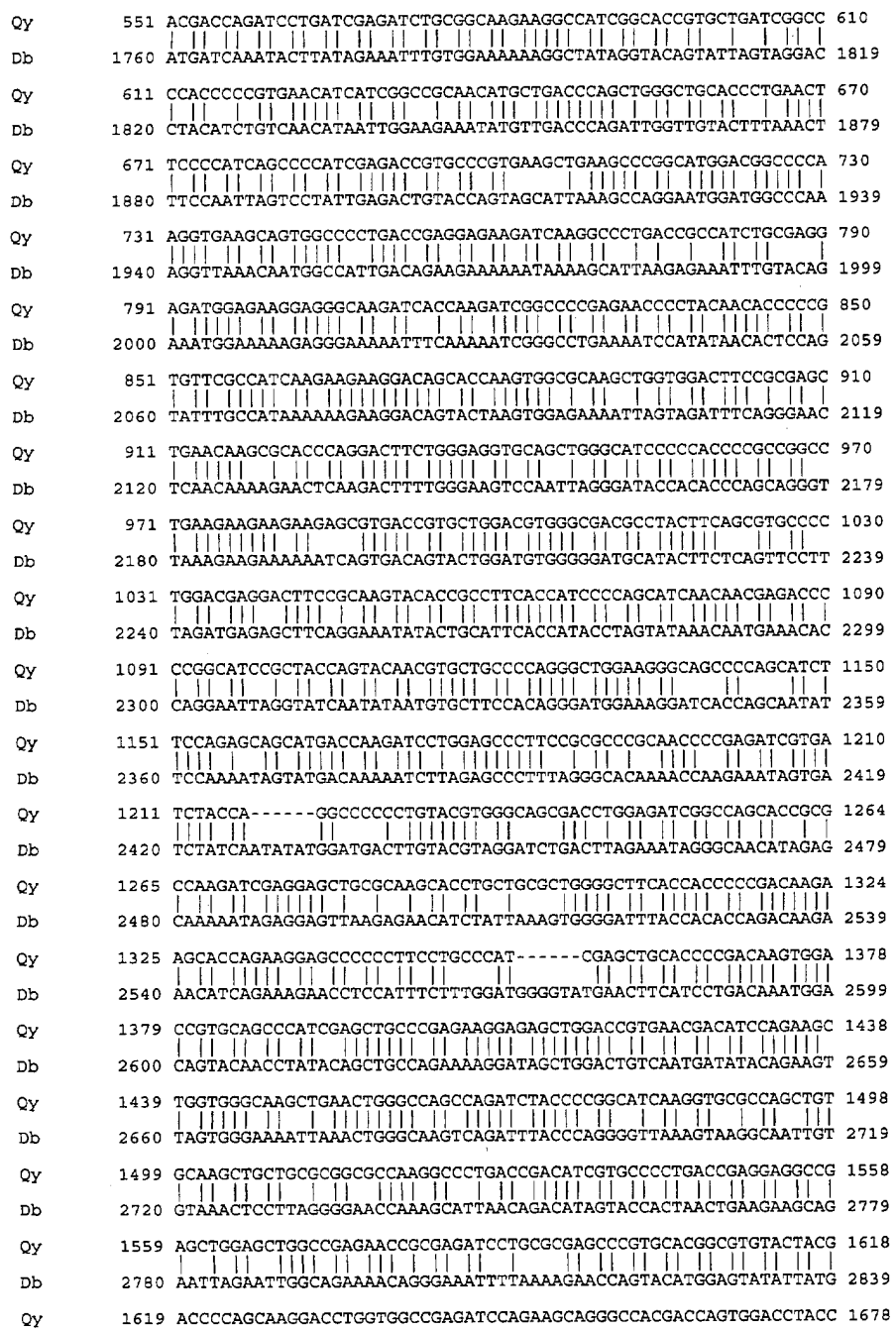
```
Qy        790 GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCC 849
Db          7 GAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCTACAACACTCCA 66

Qy        850 GTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAG 909
Db         67 GTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAG 126

Qy        910 CTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGC 969
Db        127 CTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCCGCTGGG 186

Qy        970 CTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCC 1029
Db        187 TTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCCGTTCCC 246

Qy       1030 CTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACC 1089
Db        247 TTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAACGAGACA 306

Qy       1090 CCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATC 1149
Db        307 CCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATC 366

Qy       1150 TTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG 1209
Db        367 TTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACCCAGACATCGTG 426
```

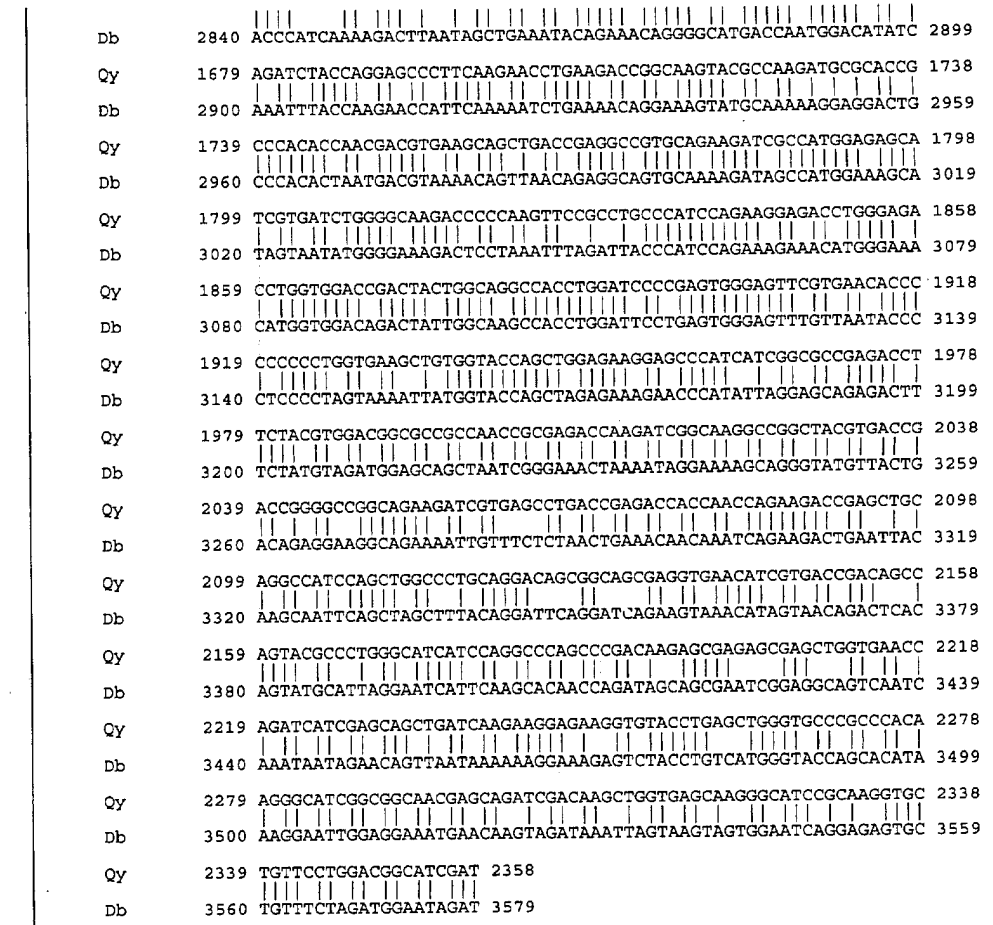




RESULT 15
US-09-184-418C-4
; Sequence 4, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 4
; LENGTH: 8992
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=92RW009; 139.1624:gag; 1690.4428:pol(N-terminus uncertain
; OTHER INFORMATION: 4373.4951:vif; 4891.5181:vpr; 5162.7801:tat; 5301.7958:rev;
; OTHER INFORMATION: 5403.5648:vpu; 5566.8148:env; 8150.8773:nef
US-09-184-418C-4

Query Match 45.5%; Score 1116.8; DB 4; Length 8992;
Best Local Similarity 68.3%; Pred. No. 2.4e-178;
Matches 1612; Conservative 0; Mismatches 732; Indels 16; Gaps 4;





Search completed: April 10, 2004, 16:22:57
Job time : 129.045 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:34:42 ; Search time 4138 Seconds
(without alignments)
17774.420 Million cell updates/sec

Title: US-09-610-313-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	86.8	3.5	2598	11	AY103647	AY103647 Zea mays
	2	85.6	3.5	951	12	BM321451	BM321451 rockefell
c	3	80.6	3.3	869	14	CK159167	CK159167 FGAS04056
	4	80.2	3.3	1132	12	BM320864	BM320864 rockefell
	5	79.6	3.2	1165	12	BM320900	BM320900 rockefell
	6	79	3.2	867	12	BM321430	BM321430 rockefell
	7	76.6	3.1	1550	12	BM321022	BM321022 rockefell
	8	75	3.0	862	12	BM321023	BM321023 rockefell
	9	74.4	3.0	914	28	BZ568300	BZ568300 pacs2-164
	10	73.2	3.0	853	12	BM321393	BM321393 rockefell
c	11	72.8	3.0	566	12	BM587428	BM587428 170006873
	12	72.6	2.9	788	14	CB643171	CB643171 OSJNEb03L
	13	72.4	2.9	753	29	CC675888	CC675888 OGWCO51TH
	14	72.2	2.9	640	10	BE601575	BE601575 HVSMEh009
	15	72.2	2.9	688	14	CB648640	CB648640 OSJNEb12C
	16	72.2	2.9	764	14	CB651670	CB651670 OSJNEb16L
	17	72.2	2.9	766	14	CB642928	CB642928 OSJNEb03F
	18	72.2	2.9	809	14	CB641397	CB641397 OSJNEb01A
	19	72.2	2.9	841	14	CB651502	CB651502 OSJNEb16H
c	20	72.2	2.9	851	29	CG260623	CG260623 OGWGE95TV
	21	71.8	2.9	1933	11	AY325173	AY325173 Rattus no
	22	71.6	2.9	500	12	BM372120	BM372120 EBro03_SQ
	23	71.6	2.9	538	12	BM368580	BM368580 EBem08_SQ
	24	71.6	2.9	540	9	AJ471121	AJ471121 AJ471121
	25	71.6	2.9	566	13	BQ464692	BQ464692 HF02P20r
	26	71.6	2.9	575	13	BU976068	BU976068 HA03B08r
	27	71.6	2.9	576	13	BU984666	BU984666 HF04K03r
	28	71.6	2.9	578	12	BM377112	BM377112 EBem05_SQ
	29	71.6	2.9	579	10	BF253640	BF253640 HVSMEf000
	30	71.6	2.9	582	12	BM372871	BM372871 EBma04_SQ
	31	71.6	2.9	586	13	BU994273	BU994273 HM06H03r
	32	71.6	2.9	588	13	BQ765498	BQ765498 EBro03_SQ
	33	71.6	2.9	596	13	BU997745	BU997745 HI08O07r
	34	71.6	2.9	600	9	AL508778	AL508778 AL508778
	35	71.6	2.9	615	13	CA009718	CA009718 HU14O18r
	36	71.6	2.9	621	13	BU986998	BU986998 HF13I08r
	37	71.6	2.9	625	13	BQ768295	BQ768295 EBro08_SQ
	38	71.6	2.9	637	13	BU999333	BU999333 HI14D01r
	39	71.6	2.9	641	13	BU978712	BU978712 HA14B02r
	40	71.6	2.9	645	13	BU977527	BU977527 HA11J06r
	41	71.6	2.9	646	13	BU968956	BU968956 HB10B12r
	42	71.6	2.9	655	13	BU996444	BU996444 HM13K04r
	43	71.6	2.9	656	12	BM370222	BM370222 EBro08_SQ
	44	71.6	2.9	659	13	CA031811	CA031811 HX11D06r
	45	71.6	2.9	660	13	BQ762385	BQ762385 EBro01_SQ

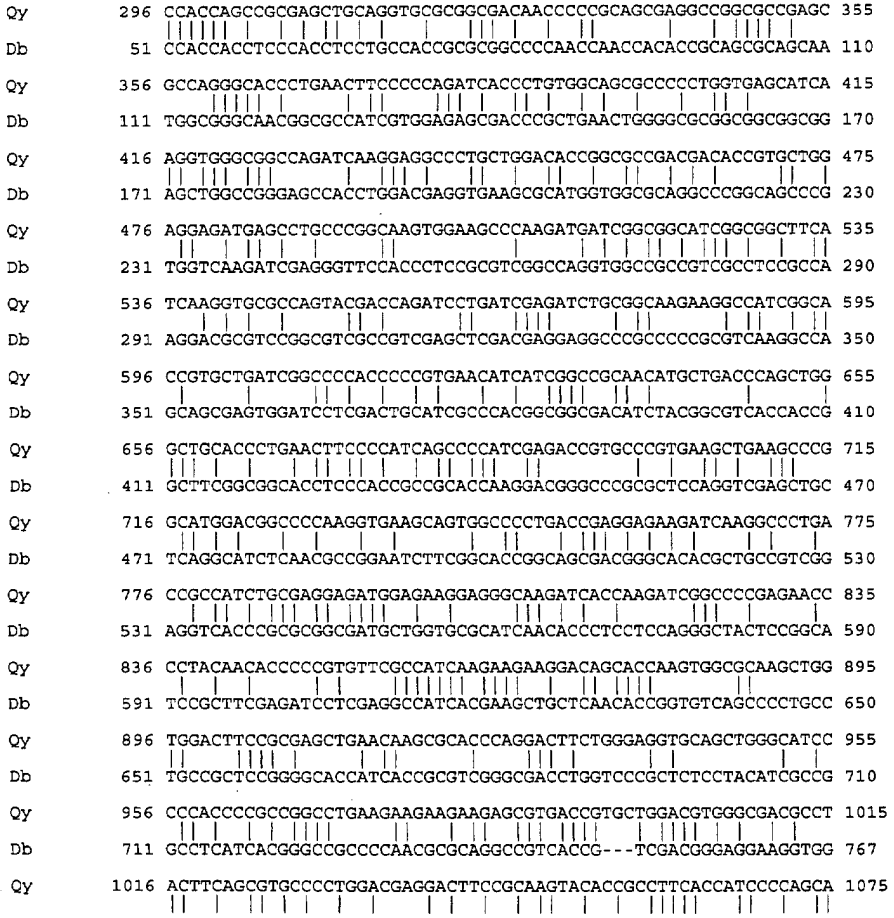
ALIGNMENTS

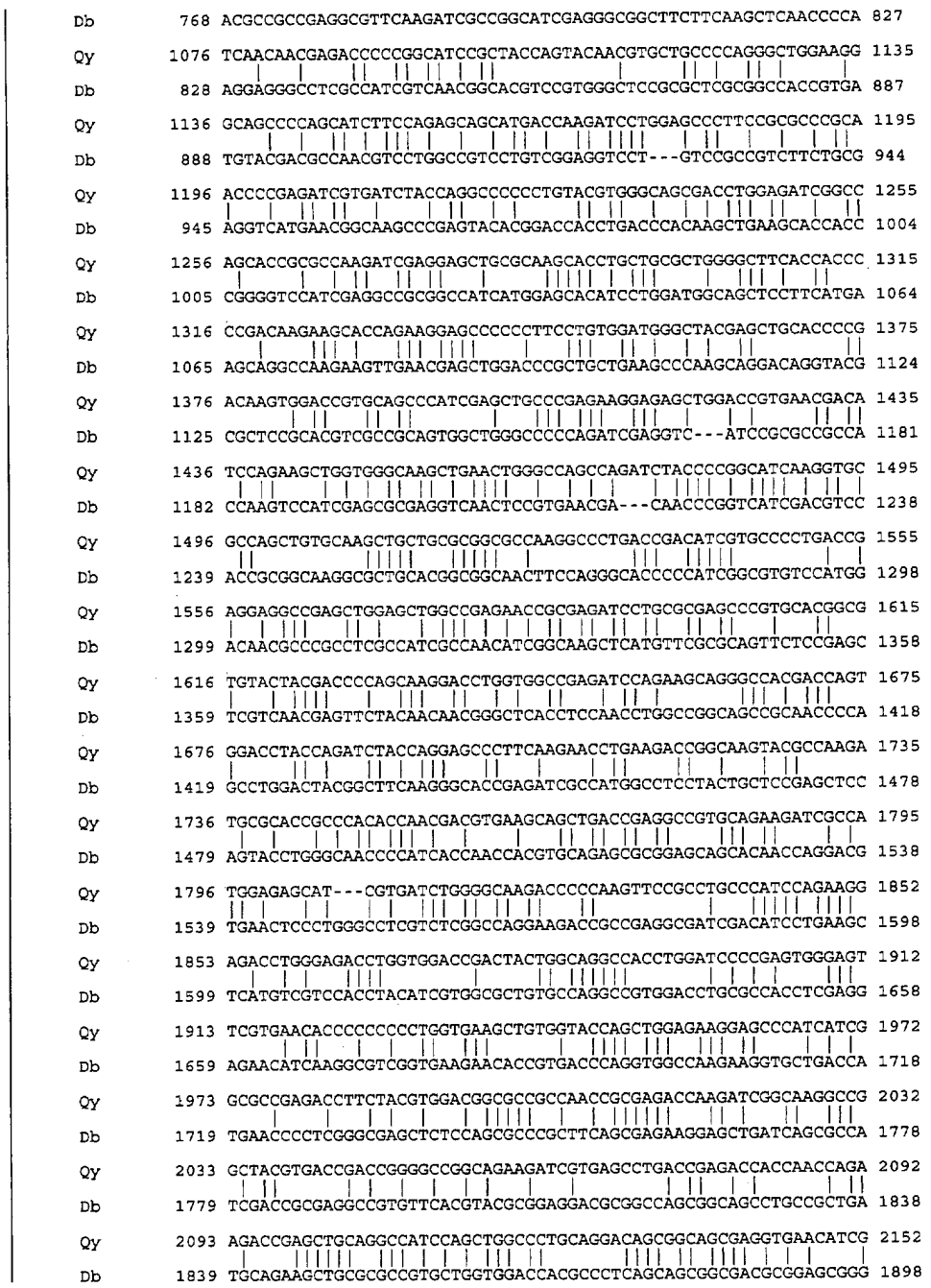
RESULT 1
AY103647
LOCUS AY103647 2598 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2598)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2598)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

```
FEATURES             Location/Qualifiers
     source          1. .2598
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638378"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN
```

Query Match 3.5%; Score 86.8; DB 11; Length 2598;
Best Local Similarity 42.1%; Pred. No. 0.59;
Matches 813; Conservative 0; Mismatches 1102; Indels 15; Gaps 5;





```
Qy       2153 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2212
Db       1899 AGCCCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGCTGCCCC 1958

Qy       2213 AGCTGGTGAA 2222
Db       1959 AGGAGGTGGA 1968
```

RESULT 2
BM321451

```
LOCUS       BM321451                951 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321451
VERSION     BM321451.1  GI:18055857
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 951)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 951   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .951
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN
```

Query Match 3.5%; Score 85.6; DB 12; Length 951;
Best Local Similarity 46.0%; Pred. No. 0.68;
Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

```
Qy       1728 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1787
Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1788 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1847
Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1848 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1907
Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1908 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1967
Db        399 CGACGGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1968 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2027
Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2028 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2087
Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2088 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2147
Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2148 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2207
Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy       2208 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2267
Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2268 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2327
Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2328 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2388 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2429
Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917
```

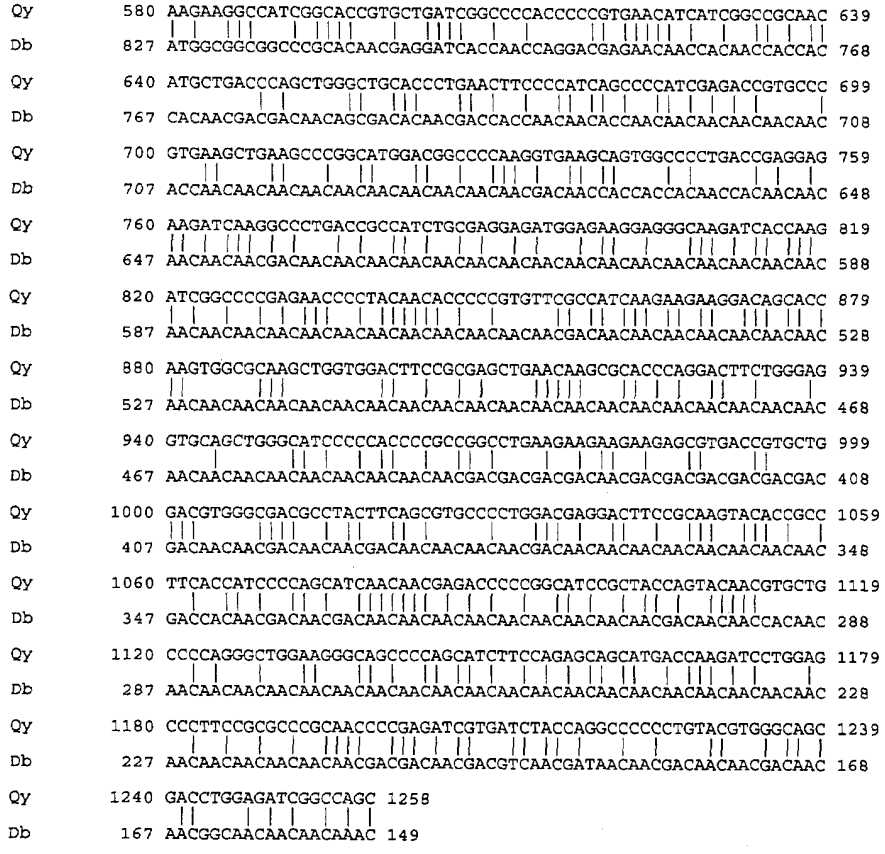
RESULT 3
CK159167/c

```
LOCUS       CK159167                869 bp    mRNA    linear   EST 05-DEC-2003
DEFINITION  FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA,
            mRNA sequence.
ACCESSION   CK159167
VERSION     CK159167.1  GI:38985053
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLt537  row: N  column: 23.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
```

(equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"

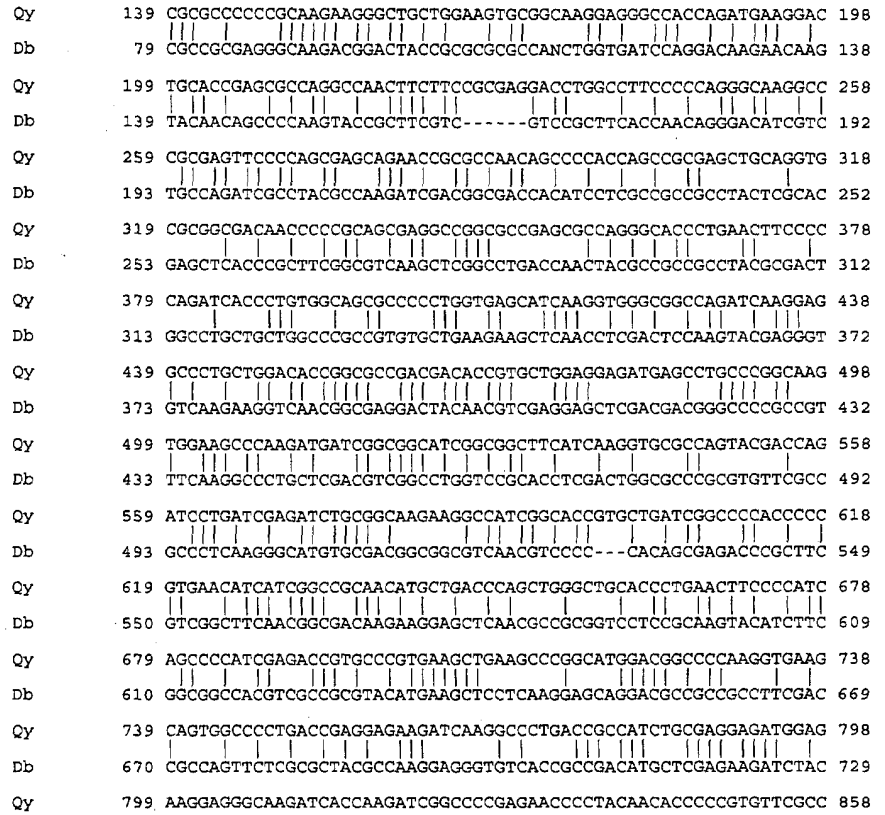
ORIGIN

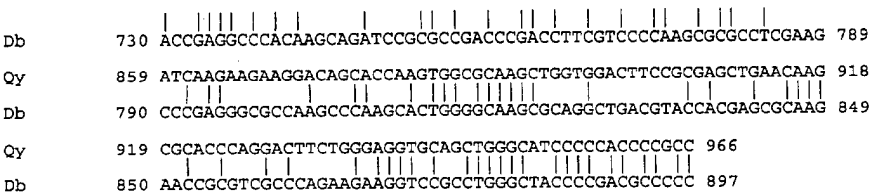
Query Match 3.3%; Score 80.6; DB 14; Length 869;
Best Local Similarity 44.9%; Pred. No. 2.7;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;



RESULT 4
BM320864
LOCUS BM320864 1132 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.
ACCESSION BM320864
VERSION BM320864.1 GI:18055270
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1132 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1132
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN

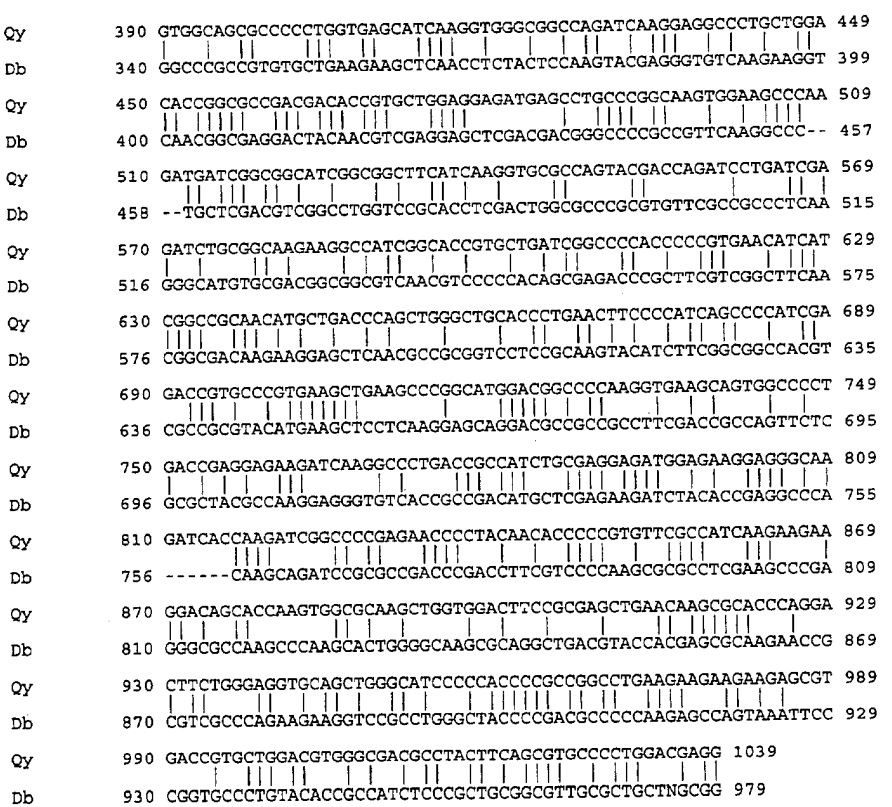
Query Match 3.3%; Score 80.2; DB 12; Length 1132;
Best Local Similarity 45.3%; Pred. No. 3.1;
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;





RESULT 5
BM320900
LOCUS BM320900 1165 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.
ACCESSION BM320900
VERSION BM320900.1 GI:18055306
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 1165)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1165 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1165
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN

Query Match 3.2%; Score 79.6; DB 12; Length 1165;
Best Local Similarity 45.1%; Pred. No. 3.7;
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;



RESULT 6
BM321430
LOCUS BM321430 867 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA sequence.
ACCESSION BM321430
VERSION BM321430.1 GI:18055836
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 867)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 867 Std Error: 0.00

```
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match              3.2%;  Score 79;  DB 12;  Length 867;
  Best Local Similarity   45.7%;  Pred. No. 4.1;
  Matches  314;  Conservative    0;  Mismatches  370;  Indels    3;  Gaps    1;

Qy         58 CAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAG 117
Db        131 CCGCACAAGATGCGCGAGTGCCTGCCGGTCATCATCCTGCTGCGCAACAGGTTGAAGTAC 190

Qy        118 GGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAG 177
Db        191 GCGCTGACCCGCCGTGAGGTCACCTCGATCGTGATGCAGCGCCTGATCAAGATCGACGGC 250

Qy        178 GAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237
Db        251 AAGGTCCGCACCGACACGCACTTCCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310

Qy        238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCC 297
Db        311 ACCGACGAGCACTTCCGCCTGCTCTACGACACCAAGGGCCGCTTCCAGGCGCACCGCATC 370

Qy        298 ACCAGCCGCGAGCTGCAGGT---GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354
Db        371 AACTCGACCGAGGCCAAGTTCAAGCTCGGCAAGGTCCGCCGCGTGCAGCTCGGCAACAAG 430

Qy        355 CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATC 414
Db        431 GGCATCCCGTACCTGGTGACCCACGACGCGCGCACGATCCGCTACCCCAACCCCGACATC 490

Qy        415 AAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTG 474
Db        491 AAGGTCAACGACACGGTCAAGATCGACCTGCGCTCGGGCAAGATCATCGACTTCGTCAGG 550

Qy        475 GAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTC 534
Db        551 TTCGAGATCGGCAACCTCGTCATGATCACTGGCGGACGCAACCTTGGCCGCGTCGGCGTC 610

Qy        535 ATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGC 594
Db        611 ATTGTGCGCCGCGAGAAGCACGAGGGCTCGTTCGAGATCATCCACGTCAAGGACGCCGTC 670

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654
Db        671 GGCCACCAGTTCGCGACGCGCCTGACCAACGTCTTCGTGATCGGCAAGGGCACCAAGTCC 730

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714
Db        731 CTCGTCACGCTGCCCGCCGGCAAGGGCATCAAGAAGTCGATCATCGAGGAGTTCCAGGCG 790

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAG 741
Db        791 CGCCACGCGCACAGGACCGAGCAGGAG 817


RESULT 7
BM321022
LOCUS       BM321022                1550 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1192 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321022
VERSION     BM321022.1  GI:18055428
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1550)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1550   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1550
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match              3.1%;  Score 76.6;  DB 12;  Length 1550;
  Best Local Similarity   45.6%;  Pred. No. 8.9;
  Matches  308;  Conservative    0;  Mismatches  364;  Indels    3;  Gaps    1;

Qy      1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db            5 CGGCACGAGGACGCAGGACCACGCGGCCGCCGCCATCGCGCAGGCCGGCGTCTCGGTCTT 64

Qy      1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db           65 CGCCTGGAAGGGCGAGAACCTCCAGGAGTACTGGGAGTGCACCTGGAAGGCCCTGTGCTT 124

Qy      1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db          125 CGGCCCCTACCAGGGCCCTCAGATCATCGTCGACGACGGCGGTGACGCGACTCTGATGAT 184

Qy      1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db          185 CCACAAGGGGTTCGCGGCCGAGGACAACCCCAAGCTGCTGGAGGACGACGAGGGCCTCGA 244

Qy      1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db          245 GGAGGTCGCCTGCCTCAACAACGTGCTCAAGCAGGTCCAGAAGGAGCAGCCCGGCTTCTG 304

Qy      2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db          305 GCACAAGATCCTCCCCGAGATCCGCGGTGTCAGCGAGGAGACGACGACTGGCGTGATGAG 364

Qy      2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db          365 GCTGTACCAGCTGCACCGCGACGGCAAGCTGCTGTTCCCGGCCGTCAACGTCAACGACTC 424

Qy      2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db          425 TG---TCACCAAGAGCAAGTTTGACAACATCTACGGCTGCCGCCACTCGCTCATCGACGG 481

Qy      2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db          482 CATCAAGCGCGGCACCGACGTGATGCTCGGCGGCAAGGTCGCCGTCGTCGCGGGCTACGG 541

Qy      2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db          542 CGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCGCGGCCAGGGCTGCCGCGTCATCGTGAC 601
```

```
Qy       2355 CGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGG 2414
Db        602 GGAGATCGACCCCATCTGCGCGCTGCAGGCGTCGATGGCGGGCTTCGAGGTCAACACGCT 661

Qy       2415 CGGCCCTAGGATCGA 2429
Db        662 CGAGGCGGGGCTCGA 676
```

RESULT 8
BM321023

```
LOCUS       BM321023                 862 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.594 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321023
VERSION     BM321023.1  GI:18055429
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 862)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 862   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .862
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN
```

```
  Query Match             3.0%;  Score 75;  DB 12;  Length 862;
  Best Local Similarity   46.2%;  Pred. No. 12;
  Matches  283;  Conservative    0;  Mismatches  327;  Indels    3;  Gaps    1;

Qy       1728 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1787
Db        220 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 279

Qy       1788 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1847
Db        280 CATCGCGCAGGCCGGAGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 339

Qy       1848 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1907
Db        340 GGAGTGCACCTNGAAGGCCCTGTGCTTCGGCCCCTACCAGGGACCTCAGANCATCGTCGA 399

Qy       1908 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1967
Db        400 CGACGGCGGTGACGCGACTCTAATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 459

Qy       1968 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2027
Db        460 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 519

Qy       2028 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2087
```

```
Db        520 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 579

Qy       2088 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2147
Db        580 CGAGGAGACGACGACTGGCGTGATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 639

Qy       2148 CATCGTGACCGACAGCCAGTACGCCCTGGGCATC---ATCCAGGCCCAGCCCGACAAGAG 2204
Db        640 GTTCCCGGCCGTCAACGTCAACGACTCTGTCACCAAGAGCAAGTTTGACAACATCTACGG 699

Qy       2205 CGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCT 2264
Db        700 CTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGGCAA 759

Qy       2265 GAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAG 2324
Db        760 GGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCGCGG 819

Qy       2325 CAAGGGCATCCGC 2337
Db        820 CCAGGGCTGCCGC 832
```

RESULT 9
BZ568300

```
LOCUS       BZ568300                 914 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_7463.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_7463, genomic survey sequence.
ACCESSION   BZ568300
VERSION     BZ568300.1  GI:27201058
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .914
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_7463"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN
```

```
  Query Match             3.0%;  Score 74.4;  DB 28;  Length 914;
  Best Local Similarity   47.0%;  Pred. No. 15;
  Matches  265;  Conservative    0;  Mismatches  296;  Indels    3;  Gaps    1;

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db        112 CTGCGCAACCAGCAGGACGGCCTGATCTTCAGCCAGGCCGGCGCCCTCGAGGTGCAGGCC 171

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db        172 GGCAGCCTGGACAACCGCCAGGGCACGCTCCAGGCCCAGGGCGACAACCAGCTGCGTATC 231
```

```
Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db        232 GGTGGCGCGCTGGACAACCAGGGCGGCCGCCTGGACAGCCGGGCCGGTAACCTAGACCTG 291

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db        292 CAGAGCGGCAGCCTCGACAACGGCGCCGGCGGCGTGCTCAACAGCGCCAAGGGTTGGCTG 351

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db        352 AAGCTGGTCACCGGGCTGTTCGACAACAGCGCCGGCGTCACCCAGGCGCAGTCGCTGGAA 411

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db        412 ATCCGCGCCGGGCAAGGCGTGCGCAACCAGCAGGGC---CATGTCTCGGCGCTGGGCGGC 468

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db        469 GACAACCGCATCGTCACCGTCGACTTCGACAACCAGGGTGGCGGCCTCTACGCCAGCGGC 528

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        529 CTGCTCAGCCTCGACGGCCAGCGCTTCCTCAACCAGGGGCGGCGGCGGGCCAGGGCGGC 588

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        589 AAGGTCGGCGCCGGGCGCATCGACTTCAGCCTGGCCGGCGCGCTGGCCAACCGCTTCGCC 648

Qy       1192 CGCAACCCCGAGATCGTGATCTAC 1215
Db        649 AGTTGGAAAGCGAGAGCGAGCTGC 672
```

RESULT 10
BM321393

LOCUS	BM321393 853 bp mRNA linear EST 03-JAN-2002
DEFINITION	rockefeller.0.1222 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.
ACCESSION	BM321393
VERSION	BM321393.1 GI:18055799
KEYWORDS	EST.
SOURCE	Mastigamoeba balamuthi
ORGANISM	Mastigamoeba balamuthi Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE	1 (bases 1 to 853)
AUTHORS	Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE	The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE	21819461
PUBMED	11830664
COMMENT	Contact: Muller Miklos Laboratory of Biochemical Parasitology The Rockefeller University 1230 York Avenue, New York, NY 10021, USA Email: mmuller@rockvax.rockefeller.edu Insert Length: 853 Std Error: 0.00 POLYA=No.

FEATURES	Location/Qualifiers
source	1. .853
	/organism="Mastigamoeba balamuthi"
	/mol_type="mRNA"
	/strain="ATCC 30984"
	/db_xref="taxon:108607"
	/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
	/note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.0%; Score 73.2; DB 12; Length 853;
Best Local Similarity 46.9%; Pred. No. 20;

Matches 299; Conservative 0; Mismatches 329; Indels 9; Gaps 2;

```
Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db        125 CTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGTGTCAAGAAG 184

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db        185 GTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCCGTTCAAGGCC 244

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        245 CTGCTCGACGTCGGCCTGGTCCGCACCTCGACCGGCGCCCGCGTGTTCGCCGCCCTCAAG 304

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        305 GGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTCGTCGGCTTC 361

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        362 AACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCAC 421

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        422 GTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGACCGCCAGTTC 481

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        482 TCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTACACCGAGGCC 541

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        542 CA------CAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAGCCC 595

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        596 GAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAGAAC 655

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db        656 CGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCCAAGAGCCAGTAAATT 715

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCG 1024
Db        716 CCCGGTGCCCTGTACACNGCCATCTCCCGCTGCGGCG 752
```

RESULT 11
BM587428/c

LOCUS	BM587428 566 bp mRNA linear EST 25-FEB-2002
DEFINITION	17000687321202 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone 19600449696110 5', mRNA sequence.
ACCESSION	BM587428
VERSION	BM587428.1 GI:18883289
KEYWORDS	EST.
SOURCE	Anopheles gambiae (African malaria mosquito)
ORGANISM	Anopheles gambiae Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE	1 (bases 1 to 566)
AUTHORS	Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE	Celera Anopheles gambiae EST project
JOURNAL	Unpublished (2002)
COMMENT	Contact: Holt R.A. Celera Genomics 45 w. Gude Dr., Rockville, MD 20850, USA Tel: 2404533151 Fax: 2404534580 Email: HoltRA@celera.com Plate: NU01004AAX row: G column: 08 Seq primer: M13 Reverse.

```
FEATURES             Location/Qualifiers
     source          1..566
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449696110"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /clone_lib="A.Gam.ad.cDNA.blood1"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                     hours  after human blood feeding. cDNA inserts >500 bp
                     cloned directionally  into pSport 1. Not 1 site is 3'.
                     Clones available through the Malaria  Research and
                     Reference Reagent Resource Center (www.malaria.mr4.org)"
ORIGIN

  Query Match             3.0%;  Score 72.8;  DB 12;  Length 566;
  Best Local Similarity   48.3%;  Pred. No. 21;
  Matches  203;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy       1950 GCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCG 2009
Db        482 GTTCGAGATCGACGCCAACGGCATCCTGCAGGTGTCGGCCGAGGACAAGGGCACGGGCAA 423

Qy       2010 CGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAG 2069
Db        422 CCGGGAAAAGATCGTCATCACCAACGACCAGAACCGCCTGACGCCGGACGACATCGAGCG 363

Qy       2070 CCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGA 2129
Db        362 CATGATCAAGGATGCGGAGCGGTTCGCCGACGACGACAAGAAGCTGAAGGAGCGCGTGGA 303

Qy       2130 CAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGC 2189
Db        302 GGCCCGCAACGAGCTCGAGAGCTACGCGTACAGCCTGAAGAACCAGCTCAGCTCGAAGGA 243

Qy       2190 CCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAA 2249
Db        242 CAAGCTGGGCGCGAGCGTGTCCGACGACGACAAGGCCAAGATGGAGGAGGCGATCGACGA 183

Qy       2250 GGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGAT 2309
Db        182 GAAGATCAAGTGGCTGGACGAGAACCAGGACACCGAGGCGGAAGAGTACAAGAAGCAGAA 123

Qy       2310 CGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGG 2369
Db        122 GAAGGAGCTGGAGGACATCGTGCAGCCCATCATTGCCAAGCTGTACGCGAGCAGTGGCGG 63


RESULT 12
CB643171
LOCUS       CB643171                 788 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb03L13.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb03L13 5', mRNA sequence.
ACCESSION   CB643171
VERSION     CB643171.1  GI:29638162
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 03  row: L  column: 13
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1..788
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb03L13"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             2.9%;  Score 72.6;  DB 14;  Length 788;
  Best Local Similarity   46.3%;  Pred. No. 24;
  Matches  314;  Conservative    0;  Mismatches  355;  Indels    9;  Gaps    2;

Qy        107 GCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGA 166
Db         99 GCGCGGCGGCGGCCGAGATGGCCGGCAGCCACCTCGACGAGGTGAAGCGCATGGTGGCGC 158

Qy        167 AGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCC 226
Db        159 AGTTCCGCGAGCCGCTGGTCAAGATCCAGGGCGCCACCCTCCGCGTCGGCCAGGTGGCCG 218

Qy        227 GCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCG 286
Db        219 CCGTCGCCCAGGCCAAGGACGCCGCCGGCGTCGCCGTCGAGCTCGACGAGGAGGCCCGCC 278

Qy        287 CCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCG 346
Db        279 CCCGCGTCAAGGCCAGCAGCGAGTGGATCCTCAACTGCATCGCCCACGGCGGCGACATCT 338

Qy        347 GCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGG 406
Db        339 ACGGCGTCACCACCGGCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGCCCCGCCC 398

Qy        407 TGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGA---CG 463
Db        399 TCCAAGTCGAGCTCCTCAGGCATCTCAACGCCGGAATCTTCGGCACTGGCTCCGATGGCC 458

Qy        464 ACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCA 523
Db        459 ACACGCTGCCGTCGGAGACGGTGCGGGCGGCCATGCTCGTGCGCATCAACACCCTCCTCC 518

Qy        524 TCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGA 583
Db        519 AGGGCTACTCCGGCATCCGGTTCGAGATCCTCGAGGCCATCACCAAGCTGCTCAACACCG 578

Qy        584 AGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGC 643
Db        579 GCGTCACGCCGTGCCTGCCGCTCCGTGGCACCATCACCGCGTCCGGTGACCTGGTTCCCC 638

Qy        644 TGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGA 703
Db        639 TGTCCTACATTGCCGGCCTCATCACCGGCCGCCCCAACGCGCAGGCCATCTCGCCCGACG 698

Qy        704 AGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGA 763
```

```
Db         699 GCAGGAAG------GTGGACGCCGCCGAGGCGTTNCAGCTCGCCGGCATCGAGGGTGGCT 752

Qy         764 TCAAGGCCCTGACCGCCA 781

Db         753 TCTTCACGCTGAACCCCA 770
```

RESULT 13
CC675888
LOCUS   CC675888   753 bp   DNA   linear   GSS 19-JUN-2003
DEFINITION OGWCO51TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0539J05, genomic survey sequence.
ACCESSION   CC675888
VERSION   CC675888.1 GI:32080584
KEYWORDS   GSS.
SOURCE   Zea mays
  ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 753)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE   Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT   Other_GSSs: OGWCO51TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1. .753
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0539J05"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

  Query Match   2.9%; Score 72.4; DB 29; Length 753;
  Best Local Similarity   46.2%; Pred. No. 25;
  Matches 281; Conservative   0; Mismatches 321; Indels   6; Gaps   1;

```
Qy          31 CAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATC 90

Db          40 CACACCACCGACGACACCTTCCGCGGGGTCCGCGTCAAGTGGACCTCCACCACGCGCACC 99

Qy          91 ATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGC 150

Db         100 GTCGACCGCAGCGGCAGCGGCGGCGGCGGTAACCCGTACAACATATTCGGCCGCGGCGGC 159

Qy         151 AAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGC 210

Db         160 GACCAGCGCGGCCTCGAGCTCCAGTTCCCGCGCCAGCACCGGGACCTCGTCCACCACCAC 219

Qy         211 CAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCC 270

Db         220 TACATCCCGCACCTCATCGACGAGGCCACCCGGATGCGGCTCAAGTCGCGGGAGCGCCGC 279

Qy         271 AGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAAC 330

Db         280 CTCTACACCAACCGCGCCACGGGGCCCTGCGACGACCACCACCGCCTCTGGACGTCGCAC 339

Qy         331 CCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTG 390

Db         340 GCCTTCGCGCACCCGTCCACGTTCGACACGCTGGCGCTCGACCCGGCGCTGCGCGACGAG 399

Qy         391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450

Db         400 GTCCGCGCCGACCTGCTGCGCTTCGCCGCGCGCCGGGACCACTACGCGCGCGTCGGCCGC 459

Qy         451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510

Db         460 GCCTGGAAGCGGGGGTACCTGCTCCACG------GCCCGCCCGGCACGGGCAAGACCAGC 513

Qy         511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570

Db         514 CTCGTCGCTGCCATCGCCAACCTGCTCGACTTCGACGTCTACGACCTGGAGCTCACCACG 573

Qy         571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630

Db         574 GTGCCCACCAACTCCCACCTCCGCCGCCTGCTCGTCTCCACCACGCCCAAGTCCGTCGTC 633

Qy         631 GGCCGCAA 638

Db         634 GTCGTCGA 641
```

RESULT 14
BE601575
LOCUS   BE601575   640 bp   mRNA   linear   EST 22-OCT-2001
DEFINITION HVSMEh0098K08f Hordeum vulgare 5-45 DAP spike EST library HVcDNA0009 (5 to 45 DAP) Hordeum vulgare subsp. vulgare cDNA clone HVSMEh0098K08f, mRNA sequence.
ACCESSION   BE601575
VERSION   BE601575.3 GI:16322423
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 640)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Close,S.J., Oates,R. and Main,D.
  TITLE   Development of a genetically and physically anchored EST resource for barley genomics: Morex 5-45 DAP spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT   On Aug 21, 2000 this sequence version replaced gi:13190104.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 461
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 534.
FEATURES   Location/Qualifiers
     source   1. .640
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEh0098K08f"
                     /tissue_type="5-45 DAP Spike"
                     /lab_host="SOLR"
                     /clone_lib="Hordeum vulgare 5-45 DAP spike EST library HVcDNA0009 (5 to 45 DAP)"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 5, 10, 15, 20, 30 and 45 DAP (Fenton). Total RNA was prepared from each pool, equal quantities of all six RNA pools were combined,

```
                poly(A) RNA was purified from the mixture, one primary
                unamplified cDNA library was made, and 1 million pfu were
                in vivo excised to give pBluescript SK(-) cDNA phagemids
                (Choi)in the TJ Close lab at the University of California,
                Riverside. Phagemids were plated and picked at the Clemson
                University Genomics Institute (CUGI) (Begum, Palmer,
                Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                sequencing and sequence analysis were performed at CUGI
                (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                The sequence has been trimmed to remove vector sequence
                and contains a minimum of 100 bases of phred value 20 or
                above. For more details on library preparation and
                sequence analysis see
                http://www.genome.clemson.edu/projects/barley. To order
                this clone see http://www.genome.clemson.edu/orders Also
                see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                Genetically and physically anchored EST resources for
                barley genomics. Barley Genetics Newsletter 31:29-30.
                (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
ORIGIN

  Query Match             2.9%;  Score 72.2;  DB 10;  Length 640;
  Best Local Similarity   47.7%;  Pred. No. 25;
  Matches  209;  Conservative    0;  Mismatches  229;  Indels    0;  Gaps    0;

Qy        789 GGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCC 848
Db        110 GGAGGAGGTGACGGTGAAGGTGTCGGCCAAGATGATCTCGGTGACGGGGCCGCGCGGCAC 169

Qy        849 CGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGA 908
Db        170 CCTGACCCGCAACTTCAAGCACCTCAACCTCGACTTCCAGCTGCAGGACGGCGGGCGCAA 229

Qy        909 GCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGG 968
Db        230 GCTCAAGGTGGACGCCTGGTTCGGCACCCGCCGCACCATGGCCGCCATCCGCACCGCCAT 289

Qy        969 CCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCC 1028
Db        290 CTCCCACGTCCAGAACCTCATCACCGGCGTCACCAAGGGCTTCCGCTACAAGATGCGCTT 349

Qy       1029 CCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGAC 1088
Db        350 CGTCTACGCTCACTTCCCCATCAACGCCTCCATCACCGCCGCCAACCGGGGCATCGAGAT 409

Qy       1089 CCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCAT 1148
Db        410 CCGCAACTTCCTCGGCGAGAAGAAGGTGAGGAAGGTGGACATGCTCGACGGGGTCACCAT 469

Qy       1149 CTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGT 1208
Db        470 CTTGCGGTCCGAGAAGGTCAAGGATGAGATCGTCCTCGACGGCAACGACATCGAGCTCGT 529

Qy       1209 GATCTACCAGGCCCCCCT 1226
Db        530 NTTCCGCTCCGGCGCCCT 547

RESULT 15
CB648640
LOCUS       CB648640                 688 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb12C01.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb12C01 5', mRNA sequence.
ACCESSION   CB648640
VERSION     CB648640.1  GI:29643633
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 12  row: C  column: 01
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1..688
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb12C01"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             2.9%;  Score 72.2;  DB 14;  Length 688;
  Best Local Similarity   46.2%;  Pred. No. 26;
  Matches  276;  Conservative    0;  Mismatches  318;  Indels    3;  Gaps    1;

Qy        107 GCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGA 166
Db         92 GCGCGGCGGCGGCCGAGATGGCCGGCAGCCACCTCGACGAGGTGAAGCGCATGGTGGCGC 151

Qy        167 AGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCC 226
Db        152 AGTTCCGCGAGCCGCTGGTCAAGATCCAGGGCGCCACCCTCCGCGTCGGCCAGGTGGCCG 211

Qy        227 GCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCG 286
Db        212 CCGTCGCCCAGGCCAAGGACGCCGCCGGCGTCGCCGTCTAGCTCGACGAGGAGGCCCGCC 271

Qy        287 CCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCG 346
Db        272 CCCGCGTCAAGGCCAGCAGCGAGTGGATCCTCAACTGCATCGCCCACGGCGGCGACATCT 331

Qy        347 GCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGG 406
Db        332 ACGGCGTCACCACCGGCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGCCCCGCCC 391

Qy        407 TGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGA---CG 463
Db        392 TCCAAGTCGAGCTCCTCAGGCATCTCAACGCCGGAATCTTCGGCACTGGCTCCGATGGCC 451

Qy        464 ACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCA 523
Db        452 ACACGCTGCCGTCGGAGACGGTGCGGGCGGCCATGCTCGTGCGCATCAACACCCTCCTCC 511

Qy        524 TCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGA 583
Db        512 AGGGCTACTCCGGCATCCGGTTCGAGATCCTCGAGGCCATCACCAAGCTGCTCAACACCG 571

Qy        584 AGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGC 643
Db        572 GCGTCACGCCGTGCCTGCCGCTCCGTGGCACCATCACCGCGTCCGGTGACCTGGTTCCCC 631
```

```
Qy        644 TGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCG 700
              || || |  | | | ||  | | | |  || |  || | |    |  | |  |||||
Db        632 TGTCCTACATTGCCGGCCTCATCACCGGCCGCCCCAACGCGCAGGCCATCTCGCCCG 688
```

Search completed: April 10, 2004, 16:16:24
Job time : 4143 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:53:16 ; Search time 6326.67 Seconds
(without alignments)
16873.640 Million cell updates/sec

Title: US-09-610-313-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2463	100.0	2463	6	AX455915	AX455915 Sequence
2	2442.2	99.2	2469	6	AX455914	AX455914 Sequence
3	2436.2	98.9	2457	6	AX455916	AX455916 Sequence
4	2046	83.1	2306	6	BD263704	BD263704 Improved
5	2046	83.1	2306	6	AR373387	AR373387 Sequence
6	2027.6	82.3	9166	6	AX427930	AX427930 Sequence
7	2025.2	82.2	2312	6	BD263706	BD263706 Improved
8	2025.2	82.2	2312	6	AR373389	AR373389 Sequence
9	2019.2	82.0	2300	6	BD263705	BD263705 Improved
10	2019.2	82.0	2300	6	AR373388	AR373388 Sequence
11	2019.2	82.0	9788	6	AX427936	AX427936 Sequence
12	2005.2	81.4	9169	6	AX427931	AX427931 Sequence
13	2001.8	81.3	9194	6	AX427926	AX427926 Sequence
14	2000.2	81.2	9194	6	AX427925	AX427925 Sequence
15	2000.2	81.2	12411	6	AX427927	AX427927 Sequence
16	1996.8	81.1	9785	6	AX427938	AX427938 Sequence
17	1993.4	80.9	9167	6	AX427933	AX427933 Sequence
18	1993.4	80.9	9170	6	AX427928	AX427928 Sequence
19	1993.4	80.9	9782	6	AX427935	AX427935 Sequence
20	1993.4	80.9	9783	6	AX427934	AX427934 Sequence
21	1993.4	80.9	9792	6	AX427932	AX427932 Sequence
22	1991.8	80.9	9189	6	AX427921	AX427921 Sequence
23	1988.8	80.7	4353	12	AF287353	AF287353 Synthetic
24	1972.2	80.1	3015	6	AX455946	AX455946 Sequence
25	1955.6	79.4	3009	6	AX455987	AX455987 Sequence
26	1942.6	78.9	4319	6	BD263639	BD263639 Improved
27	1942.6	78.9	4319	6	AR373322	AR373322 Sequence
28	1932	78.4	4352	12	AF287352	AF287352 Synthetic
29	1886.6	76.6	2577	6	AX457088	AX457088 Sequence
30	1879	76.3	4343	6	AX306428	AX306428 Sequence
31	1878.8	76.3	2305	6	BD263702	BD263702 Improved
32	1878.8	76.3	2305	6	AR373385	AR373385 Sequence
33	1875	76.1	4341	6	AX306429	AX306429 Sequence
34	1865.4	75.7	1977	6	AX455954	AX455954 Sequence
35	1857.8	75.4	1965	6	AX455952	AX455952 Sequence
36	1852.4	75.2	7897	6	AX427923	AX427923 Sequence
37	1852	75.2	2299	6	BD263703	BD263703 Improved
38	1852	75.2	2299	6	AR373386	AR373386 Sequence
39	1798.8	73.0	3024	12	AF287355	AF287355 Synthetic
40	1784	72.4	9407	6	AX427937	AX427937 Sequence
41	1651.8	67.1	4307	6	AR302563	AR302563 Sequence
42	1651.8	67.1	4307	6	AX019132	AX019132 Sequence
43	1651.8	67.1	4307	6	BD130468	BD130468 Antiviral
44	1632.6	66.3	4307	6	BD268901	BD268901 Anti-vira
45	1632.6	66.3	4307	6	AX035453	AX035453 Sequence

ALIGNMENTS

RESULT 1
AX455915
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 31 from Patent WO0204493.
ACCESSION AX455915
VERSION AX455915.1 GI:21714900
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;

CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)

FEATURES Location/Qualifiers
source 1. .2463
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YM"

ORIGIN

Query Match 100.0%; Score 2463; DB 6; Length 2463;
Best Local Similarity 100.0%; Pred. No. 5.7e-249;
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy 1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db 1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy 1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db 1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy 1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
```

```
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463

Db       2461 TTC 2463
```

RESULT 2
AX455914
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 30 from Patent WO0204493.
ACCESSION AX455914
VERSION AX455914.1 GI:21714899
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES Location/Qualifiers
source 1. .2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975(+)"
ORIGIN

Query Match 99.2%; Score 2442.2; DB 6; Length 2469;
Best Local Similarity 99.6%; Pred. No. 8.6e-247;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
```

```
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy       1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
```

```
Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Qy       2455 GGTGAATTC 2463
Db       2461 GGTGAATTC 2469
```

RESULT 3
AX455916
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 32 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES Location/Qualifiers
source 1. .2457
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YMWM"
ORIGIN

Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
Best Local Similarity 99.6%; Pred. No. 3.6e-246;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCCTGGTGAGCATCAAGGTG 420
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCCTGGTGAGCATCAAGGTG 420
Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
```

```
Qy      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy      961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db      961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy     1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db     1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy     1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db     1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy     1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db     1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy     1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db     1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy     1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db     1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy     1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db     1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy     1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db     1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy     1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db     1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy     1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db     1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy     1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db     1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy     1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db     1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy     1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db     1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy     2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db     2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy     2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db     2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy     2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db     2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy     2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db     2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy     2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db     2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy     2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db     2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy     2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
Db     2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy     2461 TTC 2463
Db     2455 TTC 2457
```

RESULT 4
BD263704
LOCUS BD263704 2306 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION BD263704
VERSION BD263704.1 GI:33073472
KEYWORDS JP 2002533124-A/71.

```
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 71 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/71
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH   Key
                       Location/Qualifiers
            FT   source          1. .2306
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             83.1%;  Score 2046;  DB 6;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 2.4e-205;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              ||||||   || || ||||| ||||| || |||||| ||| | |||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              ||||||||||||||||||  |||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||| ||      ||||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | ||||||||| |||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||||||||||||||||||| |||||||||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||||||| | ||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | |||||||||||||||||||||||||||||||||||| ||||||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              |||||||||||||||||  |||||  |||||||||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCCTGTACGTGGGCA 1237
              ||||||||||  |||||||||| ||||||||||||||||||||||||||||||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
              |||||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||||
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
              | | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
              |||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||
Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
              |||||||||||  ||  |||||||||||||||||||||||||||| |||||||||  |
```

```
Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897

Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957

Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017

Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077

Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137

Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197

Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257

Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317

Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377

Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437

Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306


RESULT 5
AR373387
LOCUS       AR373387               2306 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 82 from patent US 6602705.
ACCESSION   AR373387
VERSION     AR373387.1  GI:40075490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 82 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           83.1%;  Score 2046;  DB 6;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 2.4e-205;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
```

```
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140
Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200
Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260
Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320
Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380
Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440
Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500
Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560
Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620
Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680
Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740
Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800
Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860
Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920
Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980
Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040
Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100
Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160
Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220
```

```
Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280
Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

RESULT 6
AX427930
LOCUS AX427930 9166 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 168 from Patent WO0232943.
ACCESSION AX427930
VERSION AX427930.1 GI:21538017
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 168 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES Location/Qualifiers
source 1. .9166
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

Query Match 82.3%; Score 2027.6; DB 6; Length 9166;
Best Local Similarity 91.7%; Pred. No. 1.4e-203;
Matches 2194; Conservative 0; Mismatches 179; Indels 19; Gaps 4;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080
Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199
Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259
Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3260 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3319
Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3379
Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3380 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3439
Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3440 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3499
```

```
Qy       542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db      3500 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3559

Qy       602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db      3560 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3619

Qy       662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db      3620 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3679

Qy       722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db      3680 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3739

Qy       782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db      3740 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3799

Qy       842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db      3800 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3859

Qy       902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db      3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3919

Qy       962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db      3920 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3979

Qy      1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db      3980 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4039

Qy      1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db      4040 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4099

Qy      1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db      4100 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4159

Qy      1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db      4160 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4219

Qy      1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db      4220 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4279

Qy      1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375
Db      4280 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4339

Qy      1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435
Db      4340 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4399

Qy      1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495
Db      4400 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4459

Qy      1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
Db      4460 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4519

Qy      1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
Db      4520 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4579

Qy      1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675
Db      4580 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4639

Qy      1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735
Db      4640 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4699

Qy      1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795
Db      4700 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4759

Qy      1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855
Db      4760 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4819

Qy      1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915
Db      4820 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4879

Qy      1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975
Db      4880 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4939

Qy      1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035
Db      4940 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4999

Qy      2036 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095
Db      5000 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5059

Qy      2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155
Db      5060 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5119

Qy      2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215
Db      5120 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5179

Qy      2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275
Db      5180 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5239

Qy      2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335
Db      5240 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5299

Qy      2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db      5300 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5351
```

RESULT 7
BD263706

LOCUS	BD263706	2312 bp DNA linear PAT 17-JUL-2003
DEFINITION	Improved expression of HIV polypeptides and production of virus-like particles.	
ACCESSION	BD263706	
VERSION	BD263706.1 GI:33073474	
KEYWORDS	JP 2002533124-A/73.	
SOURCE	synthetic construct	
ORGANISM	synthetic construct artificial sequences.	
REFERENCE	1 (bases 1 to 2312)	
AUTHORS	Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.	
TITLE	Improved expression of HIV polypeptides and production of virus-like particles	
JOURNAL	Patent: JP 2002533124-A 73 08-OCT-2002; CHIRON CORP	
COMMENT	OS Artificial Sequence PN JP 2002533124-A/73	

```
PD   08-OCT-2002
PF   30-DEC-1999 JP 2000591193
PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
KARIN HARTOG,
PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
C12N5/10,
PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH   Key
                Location/Qualifiers
FT   source          1. .2312
FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 3.6e-203;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60
Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120
Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180
Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300
Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360
Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420
Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480
Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540
Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600
Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660
Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720
Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780
Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840
Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900
Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960
Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080
Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140
Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200
Qy       1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260
Qy       1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320
Qy       1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380
Qy       1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440
Qy       1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500
Qy       1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560
Qy       1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620
Qy       1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680
Qy       1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740
Qy       1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800
Qy       1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011
```

```
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071

Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131

Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191

Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251

Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311

Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371

Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 8
AR373389
LOCUS       AR373389               2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 3.6e-203;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCCAAGCACC 1291

Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351

Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411

Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
```

```
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320
Qy       1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380
Qy       1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440
Qy       1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500
Qy       1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560
Qy       1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620
Qy       1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680
Qy       1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740
Qy       1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800
Qy       1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860
Qy       2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920
Qy       2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980
Qy       2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040
Qy       2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100
Qy       2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160
Qy       2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220
Qy       2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280
Qy       2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

RESULT 9
BD263705

```
LOCUS       BD263705               2300 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263705
VERSION     BD263705.1  GI:33073473
KEYWORDS    JP 2002533124-A/72.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 72 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/72
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YMWM FH
            Key             Location/Qualifiers
            FT   source          1. .2300
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             82.0%;  Score 2019.2;  DB 6;  Length 2300;
  Best Local Similarity   93.2%;  Pred. No. 1.5e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60
Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120
Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180
Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300
Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360
Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420
Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480
```

```
Qy       638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db       481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy       698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db       541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy       758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db       601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy       818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db       661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy       878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db       721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy       938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db       781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy       998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db       841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy      1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db       901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy      1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db       961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy      1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy      1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy      1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy      1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy      1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy      1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy      1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy      1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy      1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy      1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy      1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy      1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy      1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy      1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy      2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy      2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy      2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy      2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy      2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy      2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy      2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy      2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 10
AR373388
LOCUS       AR373388               2300 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 83 from patent US 6602705.
ACCESSION   AR373388
VERSION     AR373388.1  GI:40075491
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 83 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

Query Match 82.0%; Score 2019.2; DB 6; Length 2300;
Best Local Similarity 93.2%; Pred. No. 1.5e-202;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

```
Qy   170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db     1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG  60

Qy   230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db    61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy   290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db   121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy   344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db   181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy   398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db   241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy   458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db   301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy   518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db   361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy   578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db   421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy   638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db   481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy   698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db   541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy   758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db   601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy   818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db   661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy   878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db   721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy   938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db   781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy   998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db   841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy  1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db   901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy  1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db   961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy  1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db  1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy  1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db  1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy  1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db  1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy  1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db  1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy  1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db  1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy  1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db  1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy  1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db  1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy  1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db  1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy  1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db  1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy  1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db  1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy  1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db  1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy  1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db  1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy  1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db  1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy  1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db  1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy  2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db  1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy  2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db  1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy  2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db  1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy  2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db  2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy  2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
```

```
Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377

Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437

Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300


RESULT 11
AX427936
LOCUS       AX427936               9788 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.0%;  Score 2019.2;  DB 6;  Length 9788;
  Best Local Similarity   91.6%;  Pred. No. 1e-202;
  Matches 2190;  Conservative    0;  Mismatches  178;  Indels   24;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       3141 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3194

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       3195 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3254

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364

Db       3255 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3314

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421

Db       3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481

Db       3375 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3434

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541

Db       3435 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3494

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601

Db       3495 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3554

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661

Db       3555 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3614

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721

Db       3615 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3674

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781

Db       3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3734

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841

Db       3735 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3794

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901

Db       3795 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3854

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961

Db       3855 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3914

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021

Db       3915 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3974

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081

Db       3975 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4034

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141

Db       4035 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4094

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201

Db       4095 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255

Db       4155 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4214

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315

Db       4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4274

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375

Db       4275 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4334

Qy       1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435

Db       4335 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4394

Qy       1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495

Db       4395 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4454

Qy       1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
```

```
Db   4455 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4514

Qy   1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
Db   4515 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4574

Qy   1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675
Db   4575 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4634

Qy   1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735
Db   4635 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4694

Qy   1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795
Db   4695 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4754

Qy   1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855
Db   4755 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4814

Qy   1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915
Db   4815 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4874

Qy   1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975
Db   4875 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4934

Qy   1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035
Db   4935 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4994

Qy   2036 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095
Db   4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054

Qy   2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155
Db   5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114

Qy   2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215
Db   5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

Qy   2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275
Db   5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234

Qy   2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335
Db   5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy   2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db   5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346
```

RESULT 12
AX427931
LOCUS AX427931 9169 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 169 from Patent WO0232943.
ACCESSION AX427931
VERSION AX427931.1 GI:21538018
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 169 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES Location/Qualifiers
source 1..9169
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

Query Match 81.4%; Score 2005.2; DB 6; Length 9169;
Best Local Similarity 91.4%; Pred. No. 3.1e-201;
Matches 2187; Conservative 0; Mismatches 183; Indels 22; Gaps 5;

```
Qy     14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db   2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy     74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db   3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy    134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db   3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy    194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db   3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy    254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db   3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy    314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db   3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy    365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db   3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy    422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db   3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy    482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db   3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy    542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db   3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy    602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db   3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy    662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db   3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy    722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db   3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy    782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db   3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy    842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db   3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862
```

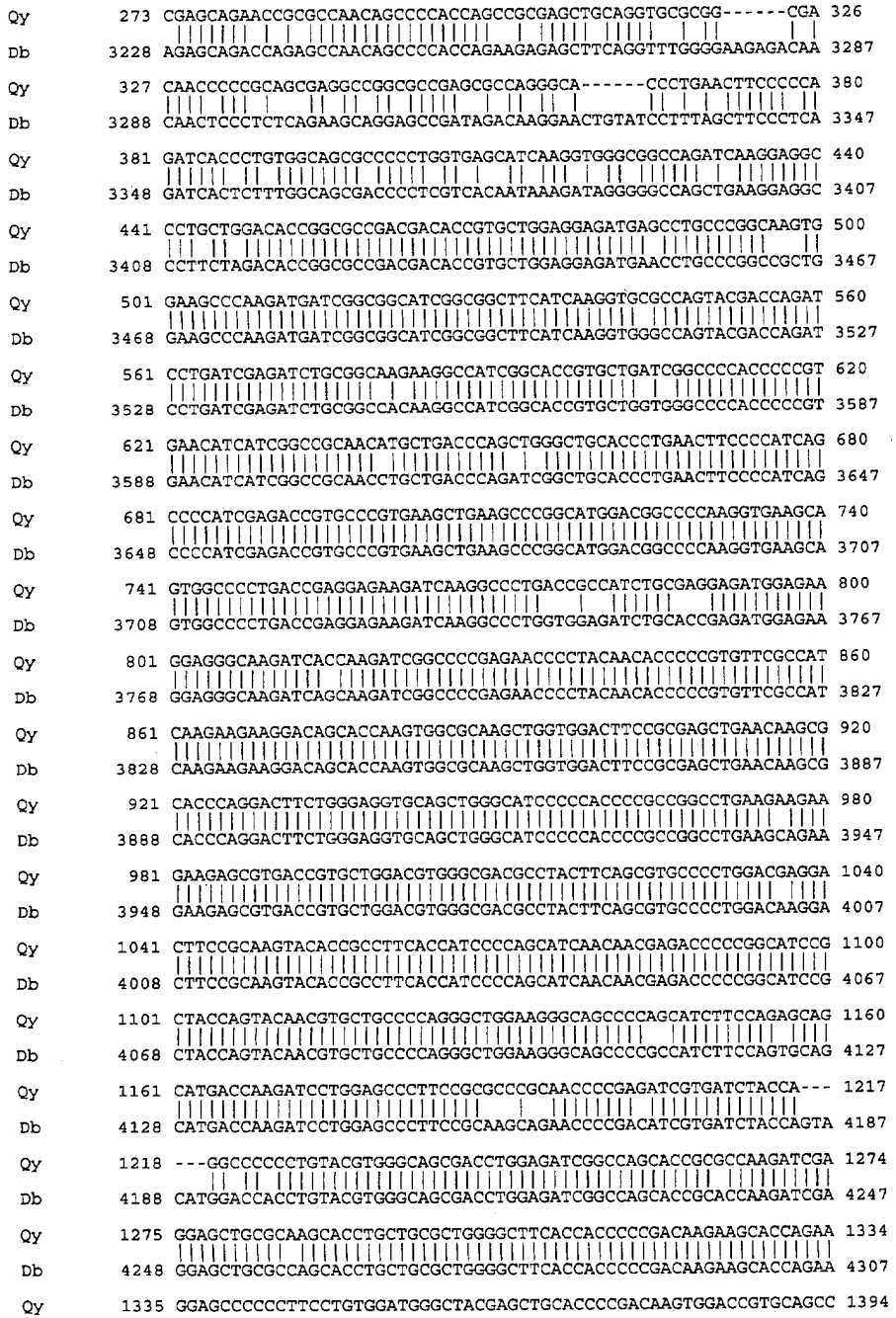
```
Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375
Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435
Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402

Qy       1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675
Db       4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy       1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735
Db       4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy       1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795
Db       4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy       1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855
Db       4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy       1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915
Db       4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy       1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975
Db       4883 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy       1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035
Db       4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy       2036 ACGTGACCGACCCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095
Db       5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy       2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155
Db       5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy       2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215
Db       5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy       2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275
Db       5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy       2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335
Db       5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy       2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db       5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354
```

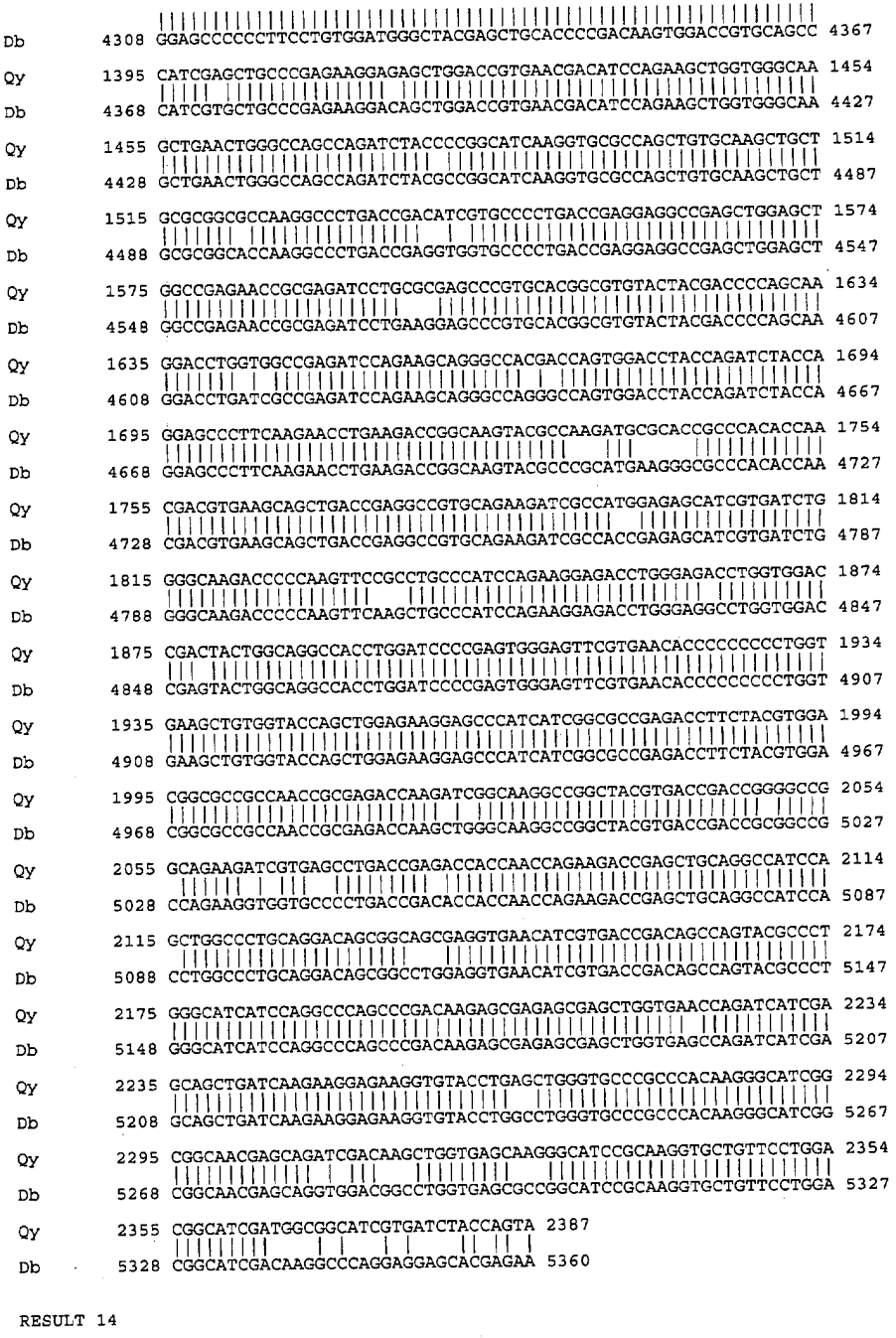
RESULT 13
AX427926

```
LOCUS       AX427926               9194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 164 from Patent WO0232943.
ACCESSION   AX427926
VERSION     AX427926.1  GI:21538013
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 164 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN
```

Query Match 81.3%; Score 2001.8; DB 6; Length 9194;
Best Local Similarity 91.3%; Pred. No. 7e-201;
Matches 2166; Conservative 0; Mismatches 188; Indels 19; Gaps 4;

```
Qy          33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db        2989 GACCAACAGCGCCACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGCAAGATCGT 3048

Qy          93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db        3049 GAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCCGCAACTGCCGCGCCCCCCGCAA 3108

Qy         153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db        3109 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGACA 3168

Qy         213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db        3169 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 3227
```





RESULT 14

AX427925
LOCUS AX427925 9194 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 163 from Patent WO0232943.
ACCESSION AX427925
VERSION AX427925.1 GI:21538012
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 163 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES Location/Qualifiers
source 1. .9194
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

Query Match 81.2%; Score 2000.2; DB 6; Length 9194;
Best Local Similarity 91.2%; Pred. No. 1e-200;
Matches 2165; Conservative 0; Mismatches 189; Indels 19; Gaps 4;

```
Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db       2989 GACCAACAGCGCCACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGCAAGATCGT 3048

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db       3049 GAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCCGCAACTGCCGCGCCCCCCGCAA 3108

Qy        153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db       3109 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGACA 3168

Qy        213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db       3169 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 3227

Qy        273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db       3228 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAA 3287

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db       3288 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCA 3347

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db       3348 GATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGGGGCCAGCTGAAGGAGGC 3407

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db       3408 CCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCCGCTG 3467

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db       3468 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGGGCCAGTACGACCAGAT 3527

Qy        561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db       3528 CCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGT 3587

Qy        621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db       3588 GAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAG 3647

Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db       3648 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 3707

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db       3708 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAA 3767

Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db       3768 GGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 3827

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db       3828 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 3887

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db       3888 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGCAGAA 3947

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db       3948 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGA 4007

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db       4008 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 4067

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db       4068 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGTGCAG 4127

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db       4128 CATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTA 4187

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db       4188 CATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGA 4247

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db       4248 GGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 4307

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db       4308 GGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 4367

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db       4368 CATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 4427

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db       4428 GCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 4487

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db       4488 GCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 4547

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db       4548 GGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 4607

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db       4608 GGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCA 4667

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db       4668 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCATGAAGGGCGCCCACACCAA 4727

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db       4728 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCACCGAGAGCATCGTGATCTG 4787
```

```
Qy   1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db   4788 GGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGAC 4847

Qy   1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db   4848 CGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 4907

Qy   1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db   4908 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 4967

Qy   1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db   4968 CGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGCGGCCG 5027

Qy   2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db   5028 CCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 5087

Qy   2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db   5088 CCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 5147

Qy   2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db   5148 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGA 5207

Qy   2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db   5208 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGG 5267

Qy   2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db   5268 CGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGGA 5327

Qy   2355 CGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db   5328 CGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

```
RESULT 15
AX427927
LOCUS       AX427927               12411 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 165 from Patent WO0232943.
ACCESSION   AX427927
VERSION     AX427927.1  GI:21538014
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 165 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..12411
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match          81.2%;  Score 2000.2;  DB 6;  Length 12411;
  Best Local Similarity   91.2%;  Pred. No. 9.6e-201;
  Matches 2165;  Conservative    0;  Mismatches  189;  Indels   19;  Gaps    4;
```

```
Qy     33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db   2989 GACCAACAGCGCCACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGCAAGATCGT 3048

Qy     93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db   3049 GAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCCGCAACTGCCGCGCCCCCCGCAA 3108

Qy    153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db   3109 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGACA 3168

Qy    213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db   3169 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 3227

Qy    273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db   3228 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAA 3287

Qy    327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db   3288 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCA 3347

Qy    381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db   3348 GATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGGGGCCAGCTGAAGGAGGC 3407

Qy    441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db   3408 CCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCCGCTG 3467

Qy    501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db   3468 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGGGCCAGTACGACCAGAT 3527

Qy    561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db   3528 CCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGT 3587

Qy    621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db   3588 GAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAG 3647

Qy    681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db   3648 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 3707

Qy    741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db   3708 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAA 3767

Qy    801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db   3768 GGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 3827

Qy    861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db   3828 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 3887

Qy    921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db   3888 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGCAGAA 3947

Qy    981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db   3948 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGA 4007

Qy   1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db   4008 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 4067

Qy   1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db   4068 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGTGCAG 4127
```

```
Qy    1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db    4128 CATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTA 4187

Qy    1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db    4188 CATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGA 4247

Qy    1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db    4248 GGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 4307

Qy    1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db    4308 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 4367

Qy    1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db    4368 CATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 4427

Qy    1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db    4428 GCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 4487

Qy    1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db    4488 GCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 4547

Qy    1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db    4548 GGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 4607

Qy    1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db    4608 GGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCA 4667

Qy    1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db    4668 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCATGAAGGGCGCCCACACCAA 4727

Qy    1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db    4728 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCACCGAGAGCATCGTGATCTG 4787

Qy    1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db    4788 GGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGAC 4847

Qy    1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db    4848 CGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 4907

Qy    1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db    4908 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 4967

Qy    1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db    4968 CGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGCGGCCG 5027

Qy    2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db    5028 CCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 5087

Qy    2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db    5088 CCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 5147

Qy    2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db    5148 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGA 5207

Qy    2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db    5208 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGG 5267

Qy    2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db    5268 CGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGGA 5327

Qy    2355 CGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db    5328 CGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

Search completed: April 10, 2004, 12:49:14
Job time : 6336.67 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:45:46 ; Search time 621 Seconds
(without alignments)
16849.133 Million cell updates/sec

Title: US-09-610-313-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2463	100.0	2463	6	ABL39960	Abl39960 Synthetic
2	2455.4	99.7	2457	7	ACA03547	Aca03547 Synthetic
3	2455.4	99.7	2457	9	ADC13265	Adc13265 DNA of HI
4	2442.2	99.2	2469	6	ABL39959	Abl39959 Synthetic
5	2436.2	98.9	2457	6	ABL39961	Abl39961 Synthetic
6	2430.2	98.7	2457	7	ACA03548	Aca03548 Synthetic
7	2430.2	98.7	2457	9	ADC13266	Adc13266 DNA of HI
8	2422.6	98.4	2445	7	ACA03546	Aca03546 Synthetic
9	2422.6	98.4	2445	9	ADC13264	Adc13264 DNA of HI
10	2415.6	98.1	3930	9	ADC13230	Adc13230 DNA of HI
11	2414	98.0	3930	9	ADC13231	Adc13231 DNA of HI
12	2414	98.0	3930	9	ADC13232	Adc13232 DNA of HI
13	2414	98.0	5184	7	ACA03591	Aca03591 Synthetic
14	2414	98.0	5184	9	ADC13279	Adc13279 DNA of HI
15	2383.6	96.8	3531	9	ADC13234	Adc13234 DNA of HI
16	2382	96.7	3537	9	ADC13236	Adc13236 DNA of HI
17	2381	96.7	5145	7	ACA03521	Aca03521 Synthetic
18	2381	96.7	5145	9	ADC13233	Adc13233 DNA of HI
19	2371	96.3	3538	9	ADC13235	Adc13235 DNA of HI
20	2322.6	94.3	3607	7	ACA03551	Aca03551 Synthetic
21	2322.6	94.3	3607	9	ADC13269	Adc13269 DNA of HI
22	2322.6	94.3	3624	7	ACA03550	Aca03550 Synthetic
23	2322.6	94.3	3624	9	ADC13268	Adc13268 DNA of HI
24	2304.4	93.6	3597	7	ACA03549	Aca03549 Synthetic
25	2304.4	93.6	3597	9	ADC13267	Adc13267 DNA of HI
26	2142	87.0	2466	7	ACA03542	Aca03542 Synthetic
27	2142	87.0	2466	7	ACC78506	Acc78506 HIV p2Pol
28	2122.8	86.2	2472	7	ACA03543	Aca03543 Synthetic
29	2122.8	86.2	2472	7	ACC78507	Acc78507 HIV p2Pol
30	2115.2	85.9	2460	7	ACA03541	Aca03541 Synthetic
31	2115.2	85.9	2460	7	ACC78505	Acc78505 HIV p2Pol
32	2114.2	85.8	3564	7	ACC78488	Acc78488 HIV GagPo
33	2114.2	85.8	3564	7	ACC78489	Acc78489 HIV GagPo
34	2113.6	85.8	4716	7	ACA03522	Aca03522 Synthetic
35	2113.6	85.8	4716	9	ADC13238	Adc13238 DNA of HI
36	2109.8	85.7	3999	7	ACC78484	Acc78484 HIV GagCo
37	2108.2	85.6	3999	7	ACC78485	Acc78485 HIV GagCo
38	2108.2	85.6	3999	7	ACC78486	Acc78486 HIV GagCo
39	2108.2	85.6	5283	7	ACA03584	Aca03584 Synthetic
40	2108.2	85.6	5283	7	ACC78529	Acc78529 HIV TatRe
41	2108	85.6	4713	7	ACA03592	Aca03592 Synthetic
42	2108	85.6	4713	9	ADC13280	Adc13280 DNA of HI
43	2107.6	85.6	3462	9	ADC13237	Adc13237 DNA of HI
44	2082.2	84.5	3735	7	ACA03545	Aca03545 Synthetic
45	2082.2	84.5	3735	7	ACC78509	Acc78509 HIV p2Pol

ALIGNMENTS

RESULT 1
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
XX
AC ABL39960;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YM SEQ ID NO:31.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 9; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene

CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 100.0%; Score 2463; DB 6; Length 2463;
Best Local Similarity 100.0%; Pred. No. 2.8e-297;
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy 1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db 1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy 1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db 1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy 1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db 1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
```

```
Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463
             |||
Db       2461 TTC 2463
```

RESULT 2
ACA03547
ID ACA03547 standard; DNA; 2457 BP.
XX
AC ACA03547;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 35; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 99.7%; Score 2455.4; DB 7; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2.4e-296;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
```

```
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 3
ADC13265
ID ADC13265 standard; DNA; 2457 BP.
XX
AC ADC13265;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 41; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 99.7%; Score 2455.4; DB 9; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2.4e-296;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
```

```
Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 4
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 99.2%; Score 2442.2; DB 6; Length 2469;
Best Local Similarity 99.6%; Pred. No. 1.1e-294;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
```

```
Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db      1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy      1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db      1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy      1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
Db      1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy      1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
Db      1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy      1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
Db      1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy      1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
Db      1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy      1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
Db      1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy      1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
Db      1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy      1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
Db      1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy      1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
Db      1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy      1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
Db      1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy      1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
Db      1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy      1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
Db      1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy      1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
Db      1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy      2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
Db      2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy      2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
Db      2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy      2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
Db      2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy      2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
Db      2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy      2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
Db      2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy      2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
Db      2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy      2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
Db      2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy      2455 GGTGAATTC 2463
Db      2461 GGTGAATTC 2469
```

RESULT 5
ABL39961
ID ABL39961 standard; DNA; 2457 BP.
XX
AC ABL39961;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX

PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 10; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
Best Local Similarity 99.6%; Pred. No. 5.9e-294;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              |||||||||||||||||||||||||||||  ||  ||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
```

```
Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463
             |||
Db       2455 TTC 2457
```

RESULT 6
ACA03548
ID ACA03548 standard; DNA; 2457 BP.
XX
AC ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 36; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 98.7%; Score 2430.2; DB 7; Length 2457;
Best Local Similarity 99.6%; Pred. No. 3.3e-293;
Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480
```

```
Qy       487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db       481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy       547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db       541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy       607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db       601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy       667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db       661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy       727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db       721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy       787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db       781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy       847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db       841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy       907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db       901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy       967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db       961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy      1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db      1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy      1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db      1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy      1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db      1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy      1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db      1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy      1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db      1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy      2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db      2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

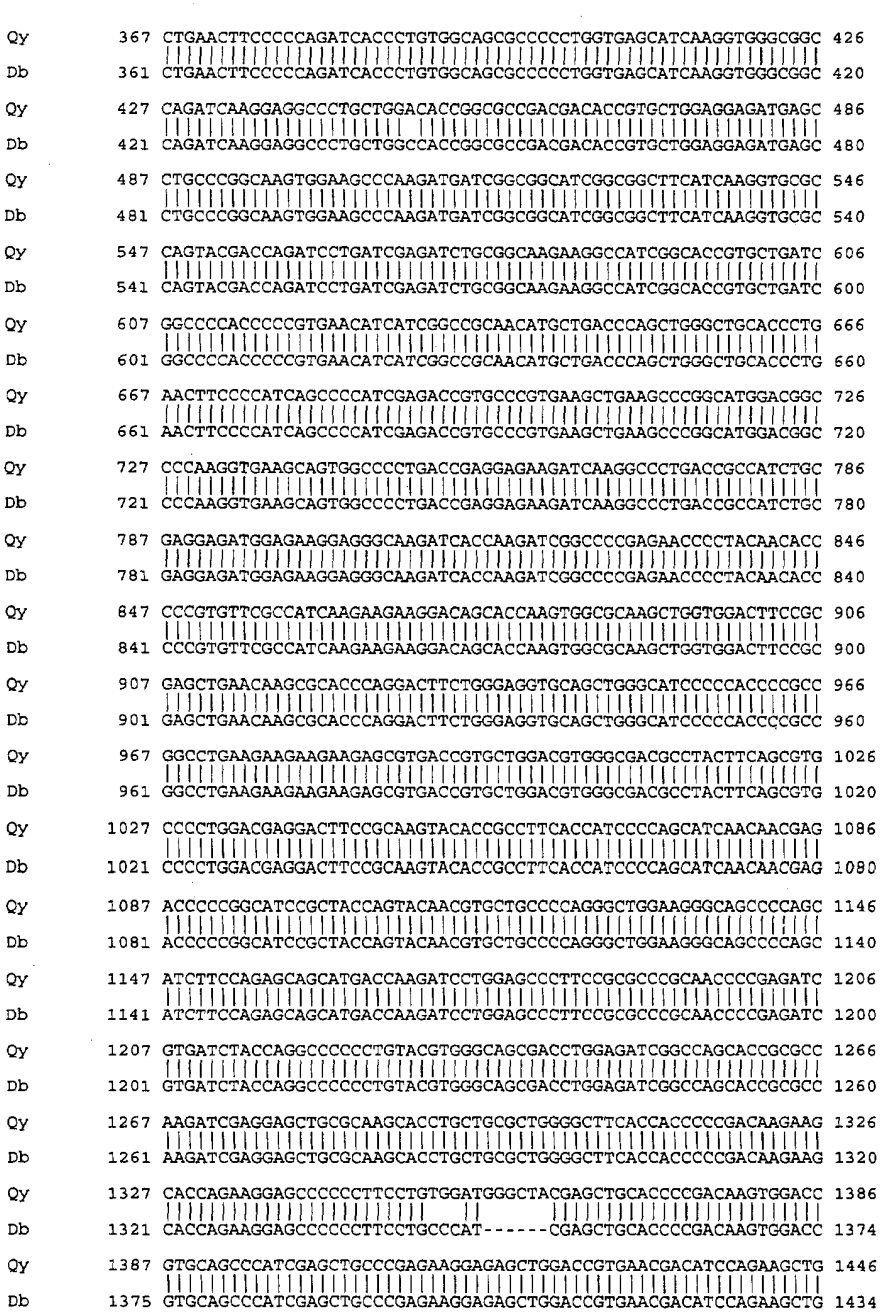
RESULT 7
ADC13266
ID ADC13266 standard; DNA; 2457 BP.
XX
AC ADC13266;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX

OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 42; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 98.7%; Score 2430.2; DB 9; Length 2457;
Best Local Similarity 99.6%; Pred. No. 3.3e-293;
Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy        7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
Db        1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy       67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db       61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy      127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
Db      121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy      187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db      181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy      247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db      241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy      307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db      301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy      367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db      361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy      427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db      421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy      487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db      481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy      547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db      541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy      607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db      601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy      667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db      661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy      727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db      721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy      787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db      781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy      847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db      841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy      907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db      901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy      967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db      961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy     1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db     1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy     1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db     1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy     1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db     1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy     1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db     1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy     1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db     1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy     1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db     1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy     1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
```

```
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 8
ACA03546
ID   ACA03546 standard; DNA; 2445 BP.
XX
AC   ACA03546;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 34; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match            98.4%;  Score 2422.6;  DB 7;  Length 2445;
  Best Local Similarity  99.6%;  Pred. No. 2.9e-292;
  Matches 2441;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360
```



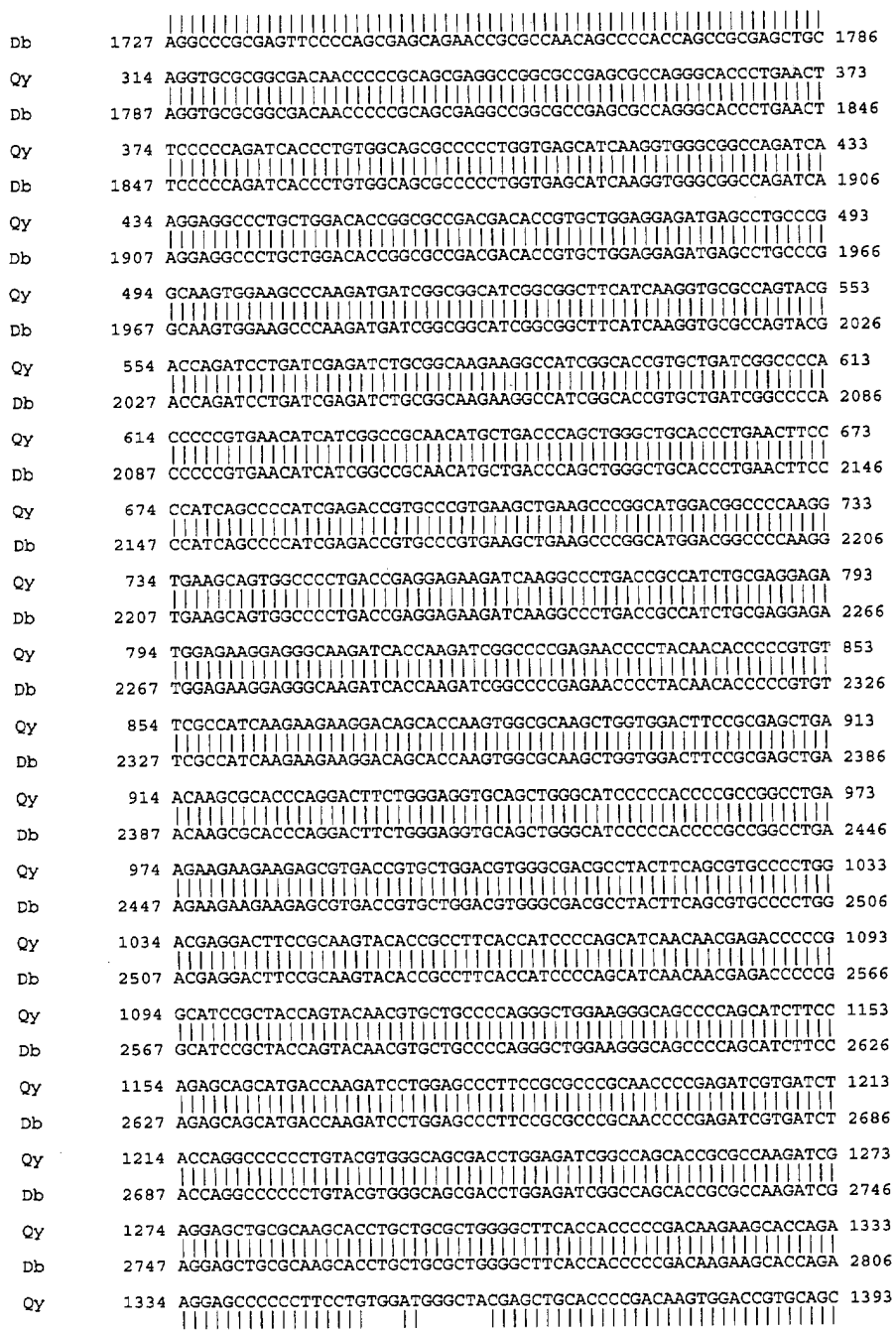
RESULT 9
ADC13264
ID ADC13264 standard; DNA; 2445 BP.
XX

```
AC   ADC13264;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 40; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match            98.4%;  Score 2422.6;  DB 9;  Length 2445;
  Best Local Similarity  99.6%;  Pred. No. 2.9e-292;
  Matches 2441;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              |||||||||||||||||||||  ||  ||||||||||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
```

```
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506

Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566

Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626

Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686

Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746

Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806

Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866

Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926

Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986

Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046

Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106

Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166

Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226

Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286

Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346

Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334

Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406

Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394

Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457

Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 10

```
ADC13230
ID   ADC13230 standard; DNA; 3930 BP.
XX
AC   ADC13230;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 6; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match            98.1%;  Score 2415.6;  DB 9;  Length 3930;
  Best Local Similarity  99.6%;  Pred. No. 2e-291;
  Matches 2434;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
```


Db 1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db 2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746
Qy 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db 2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806
Qy 1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db 2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860
Qy 1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
Db 2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920
Qy 1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db 2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980
Qy 1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db 2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040
Qy 1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db 3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100
Qy 1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db 3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160
Qy 1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db 3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220
Qy 1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db 3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280
Qy 1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
Db 3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340
Qy 1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db 3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400
Qy 1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db 3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460
Qy 1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db 3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520
Qy 2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db 3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580
Qy 2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db 3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640
Qy 2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db 3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700
Qy 2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
Db 3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760
Qy 2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
Db 3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820
Qy 2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
Db 3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880
Qy 2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db 3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 11
ADC13231
ID ADC13231 standard; DNA; 3930 BP.
XX
AC ADC13231;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagCompl1PolmutAtt_C SEQ ID NO 10.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 7; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

Query Match 98.0%; Score 2414; DB 9; Length 3930;
Best Local Similarity 99.5%; Pred. No. 3.1e-291;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```
Qy      14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy      74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy     134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy     194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

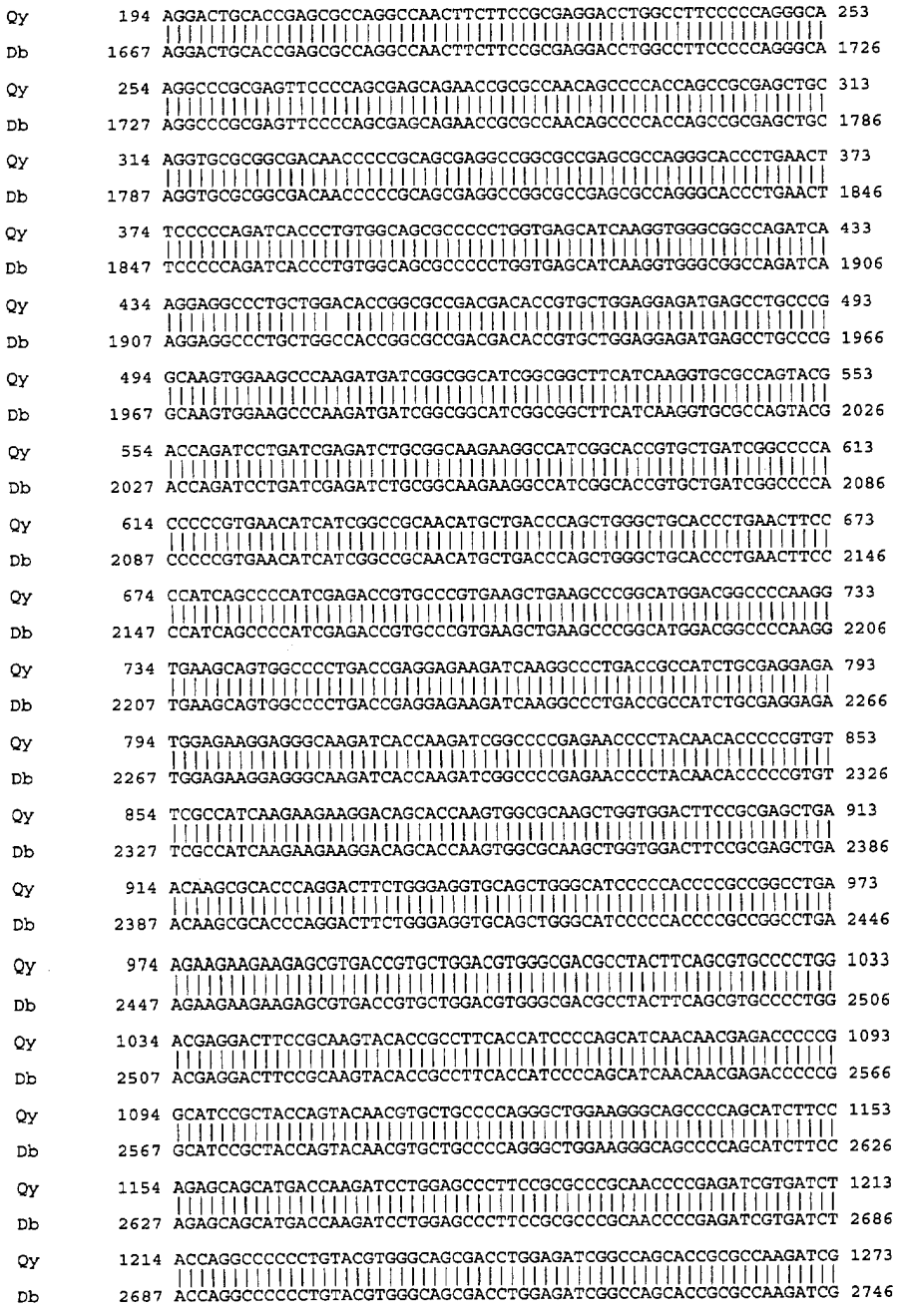
```
Db    1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy     254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy     314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy     374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy     434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy     494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy     554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy     614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy     674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy     734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy     794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy     854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy     914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy     974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy    1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy    1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy    1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy    1214 ACCAGGCCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy    1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806
```

```
Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             ||||||||||||||||  ||      ||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 12
ADC13232
ID ADC13232 standard; DNA; 3930 BP.
XX
AC ADC13232;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagCompl1PolmutIna_C SEQ ID NO 11.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 8; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 98.0%; Score 2414; DB 9; Length 3930;
Best Local Similarity 99.5%; Pred. No. 3.1e-291;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
```


Qy 194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db 1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db 2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746
Qy 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db 2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806
Qy 1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db 2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860
Qy 1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
Db 2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920
Qy 1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db 2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980
Qy 1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db 2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040
Qy 1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db 3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100
Qy 1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db 3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160
Qy 1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db 3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220
Qy 1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db 3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280
Qy 1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
Db 3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340
Qy 1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db 3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400
Qy 1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db 3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460
Qy 1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db 3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520
Qy 2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db 3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580
Qy 2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db 3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640
Qy 2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db 3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700
Qy 2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
Db 3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760
Qy 2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
Db 3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820
Qy 2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413

```
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 13
ACA03591
ID   ACA03591 standard; DNA; 5184 BP.
XX
AC   ACA03591;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #74.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 79; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

  Query Match           98.0%;  Score 2414;  DB 7;  Length 5184;
  Best Local Similarity  99.5%;  Pred. No. 3e-291;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
```

```
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000
Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060
Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             |||||||||||||||||  ||      ||||||||||||||||||||||||||||||||
Db       4061 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 4114
Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4115 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 4174
Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4175 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 4234
Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4235 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 4294
Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4295 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 4354
Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4355 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 4414
Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4415 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 4474
Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4475 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 4534
Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4535 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 4594
Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4595 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 4654
Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4655 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 4714
Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4715 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 4774
Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4775 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 4834
Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4835 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 4894
Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4895 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 4954
Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4955 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 5014
```

```
Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5015 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 5074
Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5075 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 5134
Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||
Db       5135 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 14
ADC13279
ID ADC13279 standard; DNA; 5184 BP.
XX
AC ADC13279;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct TatRevNefgagCpolIna_C SEQ ID NO 58.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 55; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

  Query Match   98.0%; Score 2414; DB 9; Length 5184;
  Best Local Similarity 99.5%; Pred. No. 3e-291;
  Matches 2433; Conservative   0; Mismatches   5; Indels   6; Gaps   1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800
Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db      2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860
Qy       134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db      2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920
Qy       194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db      2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980
Qy       254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db      2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040
Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db      3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100
Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db      3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160
Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db      3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220
Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db      3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280
Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db      3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340
Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db      3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400
Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db      3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460
Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db      3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520
Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580
Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640
Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700
Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820
Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880
```

```
Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000
Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060
Qy      1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db      4061 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 4114
Qy      1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
Db      4115 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 4174
Qy      1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db      4175 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 4234
Qy      1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db      4235 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 4294
Qy      1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db      4295 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 4354
Qy      1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db      4355 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 4414
Qy      1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db      4415 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 4474
Qy      1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db      4475 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 4534
Qy      1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
Db      4535 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 4594
Qy      1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db      4595 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 4654
Qy      1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db      4655 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 4714
Qy      1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db      4715 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 4774
Qy      2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db      4775 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 4834
Qy      2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db      4835 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 4894
Qy      2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db      4895 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 4954
```

```
Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4955 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 5014

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5015 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 5074

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5075 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 5134

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||
Db       5135 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 15
ADC13234
ID ADC13234 standard; DNA; 3531 BP.
XX
AC ADC13234;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagPolmut_C SEQ ID NO 13.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 10; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3531 BP; 802 A; 1210 C; 1096 G; 423 T; 0 U; 0 Other;

Query Match 96.8%; Score 2383.6; DB 9; Length 3531;
Best Local Similarity 98.8%; Pred. No. 1.9e-287;
Matches 2414; Conservative 0; Mismatches 24; Indels 6; Gaps 1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| |||||||||||||| || |  |  ||  ||  | || ||| |      |
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2048 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2107

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2108 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2167
```

```
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2168 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2227

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2288 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2347

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2348 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2407

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             |||||||||||||||||  ||      ||||||||||||||||||||||||||||||||||||
Db       2408 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2461

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2462 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2521

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2522 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2581

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2582 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 2641

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2642 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 2701

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2702 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 2761

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2762 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 2821

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2822 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 2881

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2882 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 2941

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2942 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3001

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3002 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3061

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3062 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3121

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3122 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3181

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3182 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3241

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3242 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3301

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3302 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCCCACAAGGGCATCG 3361

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3362 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3421

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3422 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3481

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||
Db       3482 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3525
```

Search completed: April 10, 2004, 07:32:28
Job time : 629 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:48:32 ; Search time 118.333 Seconds
(without alignments)
11550.795 Million cell updates/sec

Title: US-09-610-313-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

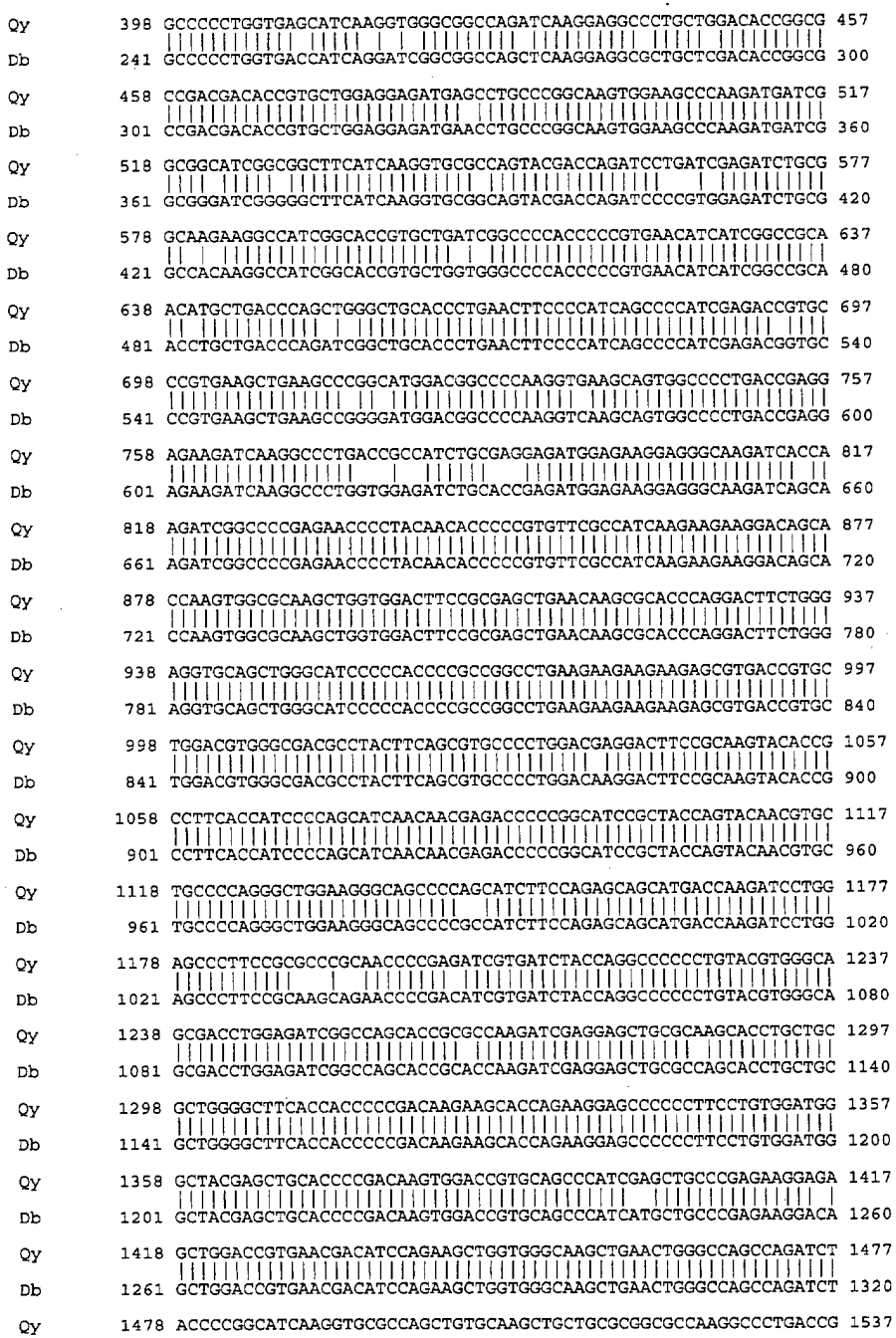
Result No.	Score	% Query Match	Length	DB	ID	Description
1	2046	83.1	2306	4	US-09-475-515-82	Sequence 82, Appl
2	2025.2	82.2	2312	4	US-09-475-515-84	Sequence 84, Appl
3	2019.2	82.0	2300	4	US-09-475-515-83	Sequence 83, Appl
4	1942.6	78.9	4319	4	US-09-475-515-6	Sequence 6, Appli
5	1878.8	76.3	2305	4	US-09-475-515-80	Sequence 80, Appl
6	1852	75.2	2299	4	US-09-475-515-81	Sequence 81, Appl
7	1651.8	67.1	4307	4	US-09-552-950-2	Sequence 2, Appli
8	1624.6	66.0	9772	4	US-09-552-950-5	Sequence 5, Appli
9	1566	63.6	8366	4	US-09-872-733A-6	Sequence 6, Appli
10	1530.6	62.1	4338	4	US-09-872-733A-1	Sequence 1, Appli
11	1203.6	48.9	9010	4	US-09-184-418C-8	Sequence 8, Appli
12	1172.6	47.6	8972	4	US-09-184-418C-9	Sequence 9, Appli
13	1165.4	47.3	8959	4	US-09-184-418C-11	Sequence 11, Appl
14	1163.8	47.3	2467	4	US-09-872-733A-3	Sequence 3, Appli
15	1142	46.4	8992	4	US-09-184-418C-4	Sequence 4, Appli
16	1106.8	44.9	2601	3	US-09-117-217-7	Sequence 7, Appli
17	1106.8	44.9	2601	3	US-09-117-217-9	Sequence 9, Appli
18	1106.8	44.9	2601	3	US-09-117-217-11	Sequence 11, Appl
19	1106.8	44.9	2601	3	US-09-117-217-13	Sequence 13, Appl
20	1106.8	44.9	2601	4	US-09-735-487-7	Sequence 7, Appli
21	1106.8	44.9	2601	4	US-09-735-487-9	Sequence 9, Appli
22	1106.8	44.9	2601	4	US-09-735-487-11	Sequence 11, Appl
23	1106.8	44.9	2601	4	US-09-735-487-13	Sequence 13, Appl
24	1106.8	44.9	4307	4	US-09-552-950-1	Sequence 1, Appli
25	1106.8	44.9	9719	4	US-09-700-304-1	Sequence 1, Appli
26	1105.2	44.9	9050	4	US-09-184-418C-7	Sequence 7, Appli
27	1100.4	44.7	7399	2	US-08-418-848A-9	Sequence 9, Appli
28	1100.4	44.7	9709	2	US-08-188-583-5	Sequence 5, Appli
29	1100.4	44.7	9709	3	US-08-388-353-1	Sequence 1, Appli
30	1100.4	44.7	9709	3	US-08-488-551B-1	Sequence 1, Appli
31	1100.4	44.7	9709	4	US-09-309-572-15	Sequence 15, Appl
32	1100.4	44.7	9709	4	US-09-718-096-15	Sequence 15, Appl
33	1100.4	44.7	12479	4	US-09-318-138-13	Sequence 13, Appl
34	1100.4	44.7	12494	3	US-08-935-312-13	Sequence 13, Appl
35	1100.4	44.7	12494	3	US-08-848-760B-33	Sequence 33, Appl
36	1100.4	44.7	15581	3	US-08-646-538-35	Sequence 35, Appl
37	1100.4	44.7	15581	3	US-09-503-222-35	Sequence 35, Appl
38	1099.2	44.6	3000	4	US-09-184-418C-74	Sequence 74, Appl
39	1095.8	44.5	8968	4	US-09-184-418C-1	Sequence 1, Appli
40	1090.8	44.3	4313	4	US-09-475-515-3	Sequence 3, Appli
41	1090.8	44.3	9737	2	US-08-944-449-7	Sequence 7, Appli
42	1090.8	44.3	9737	4	US-09-353-362-7	Sequence 7, Appli
43	1082.8	44.0	8954	4	US-09-184-418C-6	Sequence 6, Appli
44	1077.6	43.8	3017	4	US-09-184-418C-83	Sequence 83, Appl
45	1075.4	43.7	3011	4	US-09-184-418C-101	Sequence 101, App

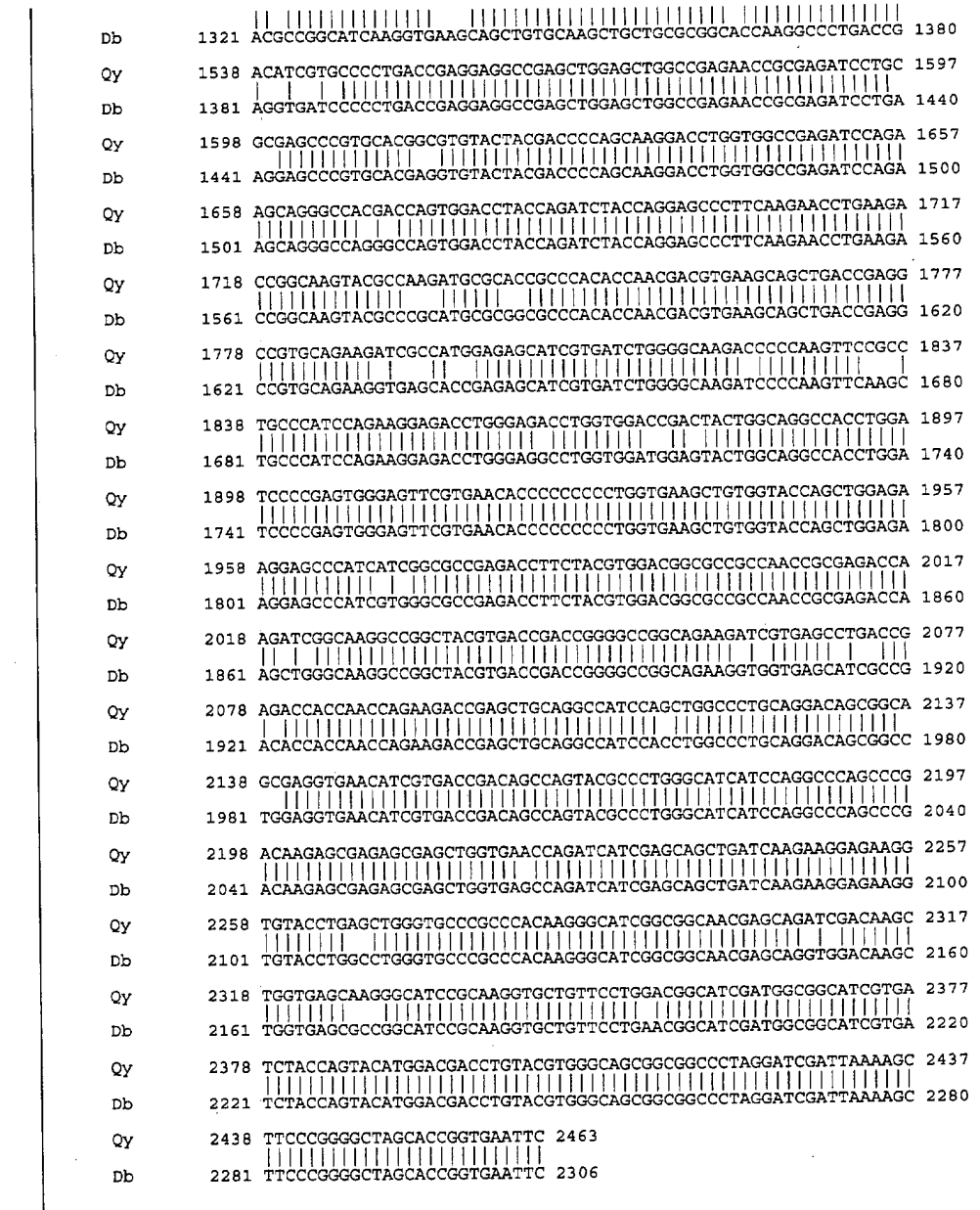
ALIGNMENTS

RESULT 1
US-09-475-515-82
; Sequence 82, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 2306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

Query Match 83.1%; Score 2046; DB 4; Length 2306;
Best Local Similarity 93.6%; Pred. No. 0;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

```
Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
            |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
            ||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db       61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
            |||||||||| |||||||||||||||| | ||||      |||||  ||  |||||||
Db      121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy      344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
            |||||||||| |||||||||||| ||      ||||||||||||||||||||||||||||
Db      181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
```



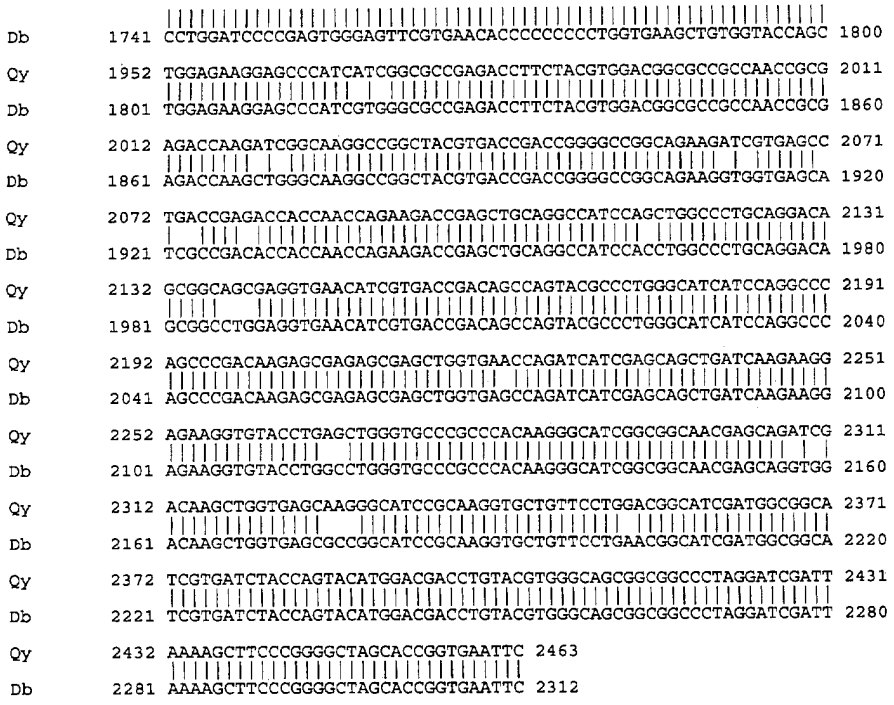


RESULT 2
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 84
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt(+)
US-09-475-515-84

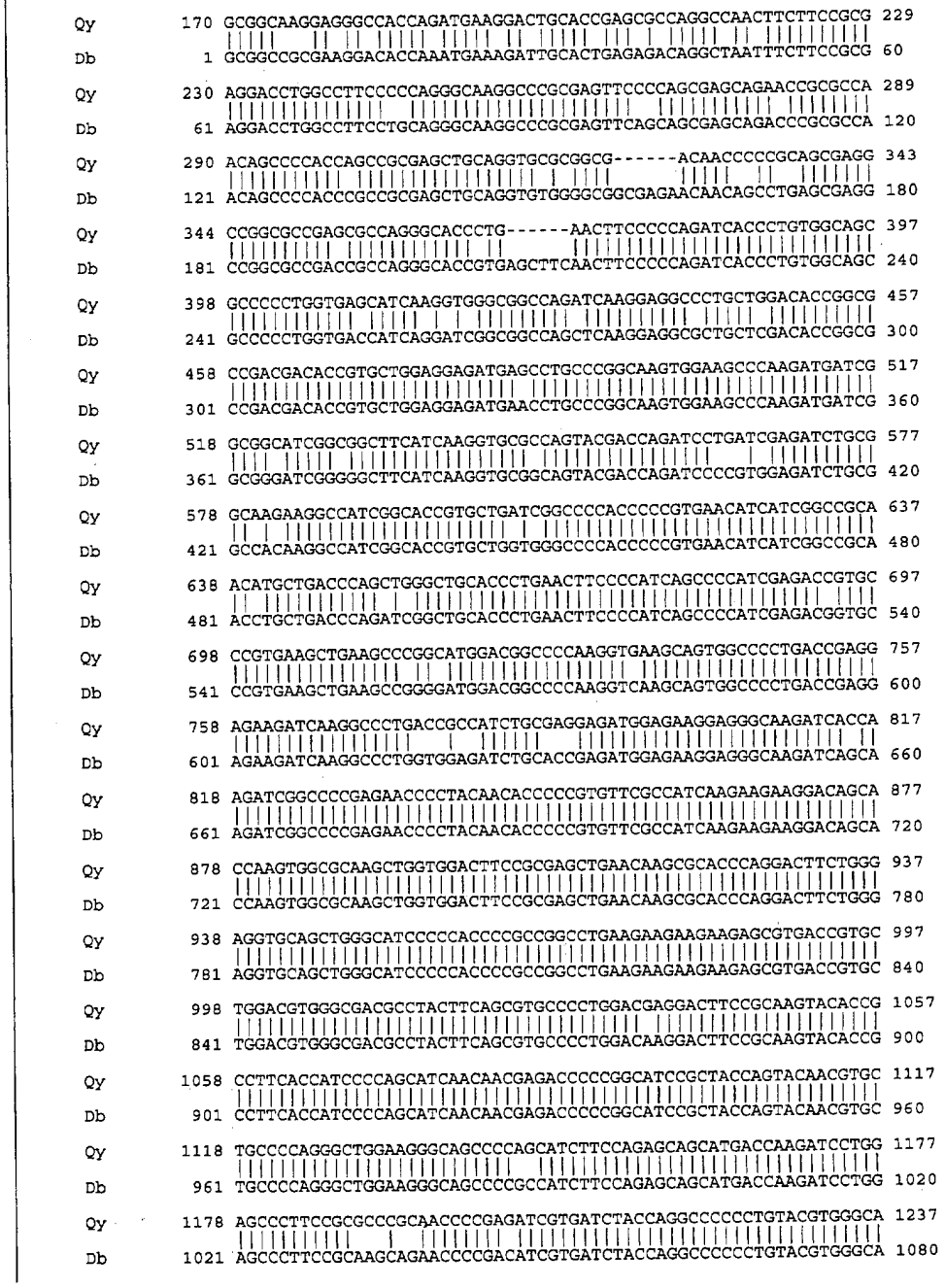
Query Match 82.2%; Score 2025.2; DB 4; Length 2312;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

```
Qy   170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db     1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG  60

Qy   230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db    61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy   290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db   121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy   344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db   181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy   398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db   241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy   458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db   301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy   518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db   361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy   578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db   421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy   638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db   481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy   698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db   541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy   758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db   601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy   818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db   661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy   878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db   721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy   938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db   781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy   998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db   841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy  1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db   901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy  1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db   961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy  1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db  1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy  1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db  1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy  1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db  1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy  1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db  1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy  1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db  1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy  1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db  1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy  1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db  1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy  1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db  1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy  1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db  1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy  1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db  1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy  1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db  1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy  1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db  1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy  1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
```



RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YMWM
US-09-475-515-83

Query Match 82.0%; Score 2019.2; DB 4; Length 2300;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

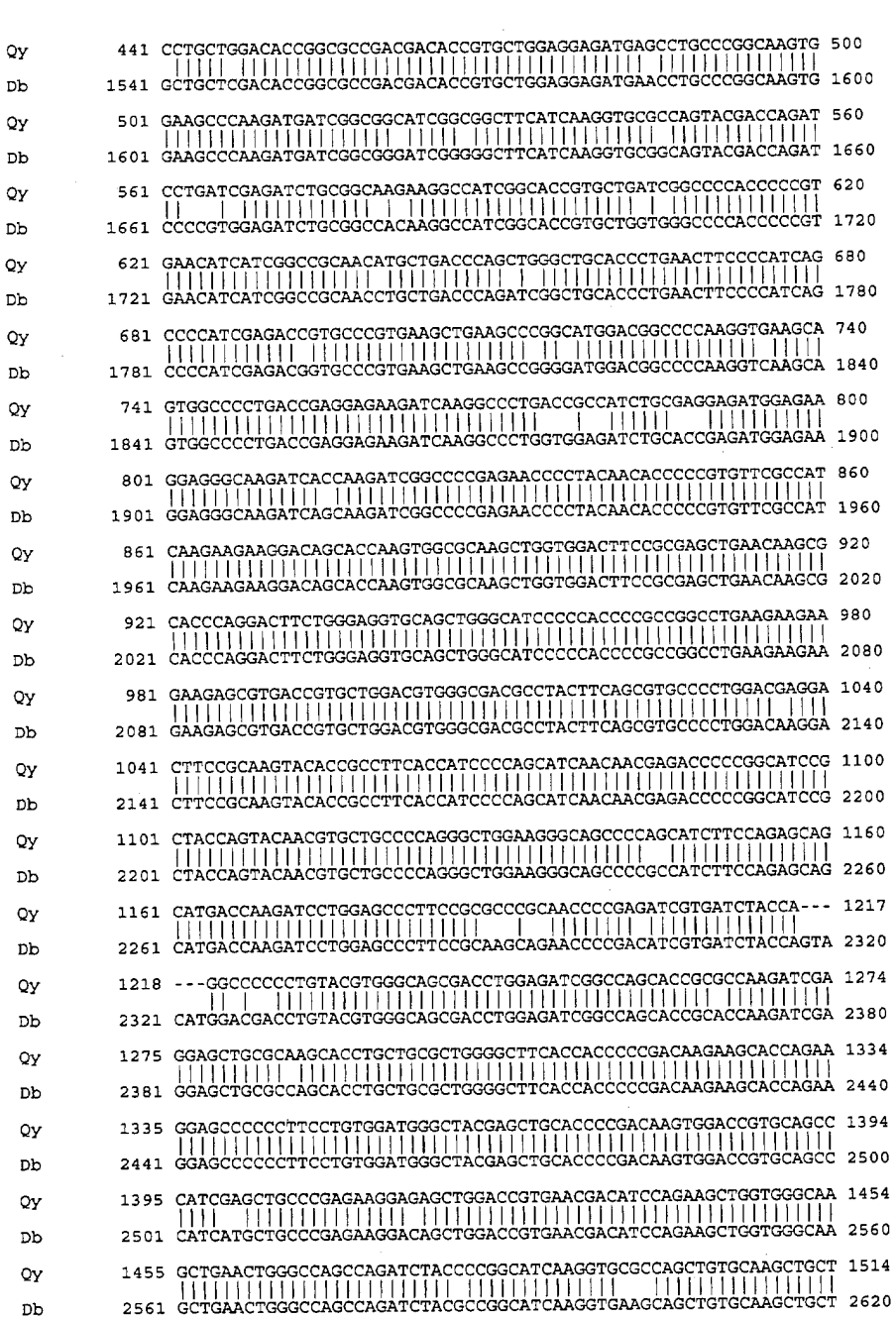


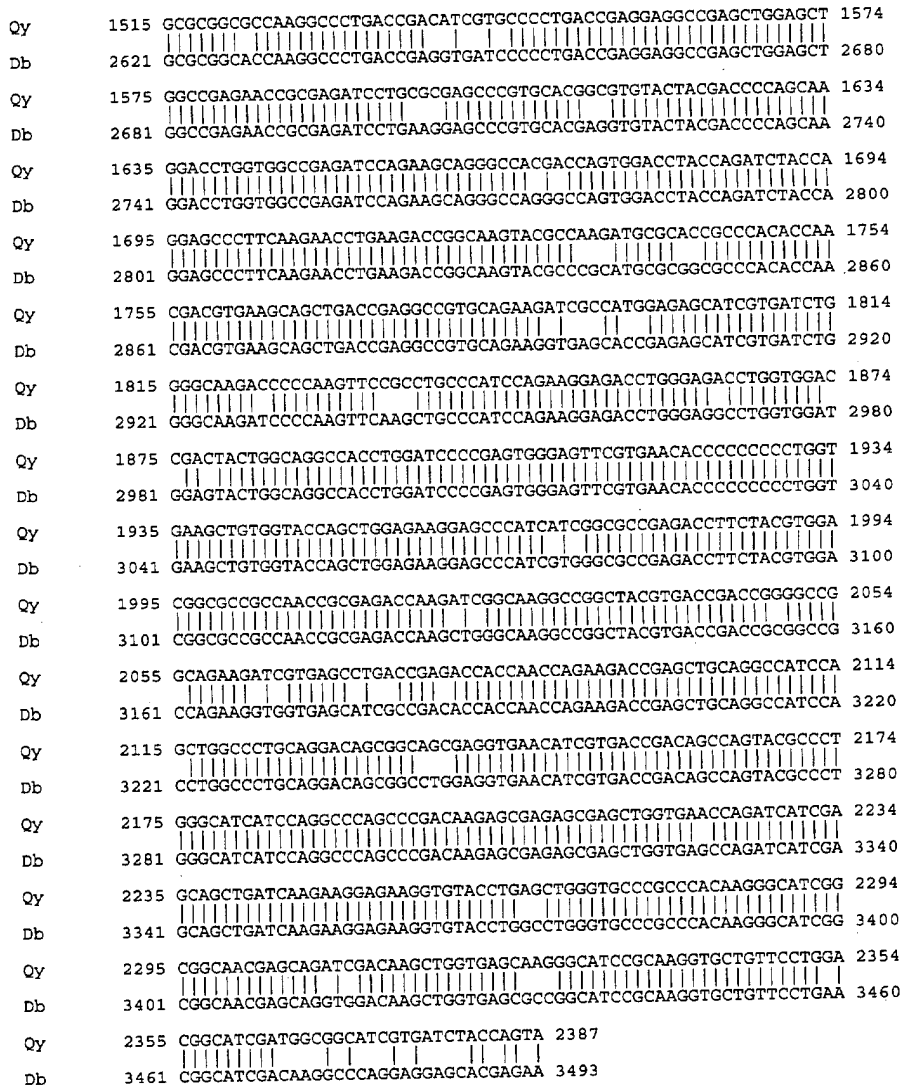
```
Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
             ||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||||||
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCCTTCCTGTGGATGG 1357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||    ||
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCCTTCCTGCCCAT-- 1198

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
                 |||||||||||||||||||||||||||||||||||||   |||||||||||||||| |
Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
             ||  ||||||||||||||  ||||||||||||||||||||||||| ||||||||||||||
Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
             |   |  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
                 |||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
             |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
             |||||||||||||   ||||||  ||||||||||||||||||||||||||||||||||||
Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
             |||||||||||  |   ||  |||||||||||||||||||||||| ||||||||||  |
Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
             |||||||||||||||||||||||||| |||||||||  || ||||||||||||||||||||
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
             ||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
             || | |||||||||||||||||||||||||||||||||||||||| | |||||| |  |||
Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
             | |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
             |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
             |||||||| |||||||||||||||||||||||||||||||||||||||| | |||||||
Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
             |||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
             ||||||||||||||||||||||||||
Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

RESULT 4
US-09-475-515-6
; Sequence 6, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: HIV-Gag-polymerase
US-09-475-515-6

Query Match 78.9%; Score 1942.6; DB 4; Length 4319;
Best Local Similarity 89.7%; Pred. No. 0;
Matches 2129; Conservative 0; Mismatches 225; Indels 19; Gaps 4;

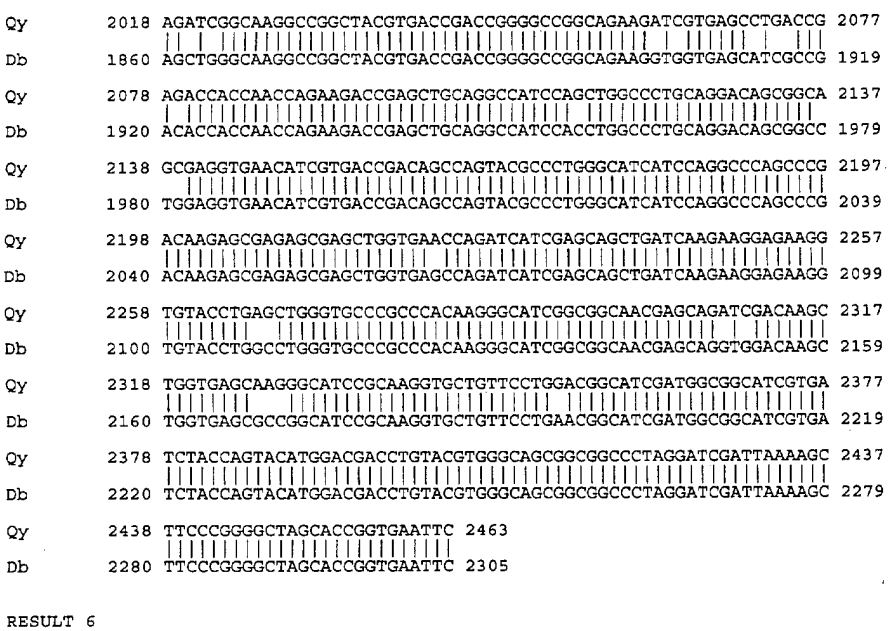
```
Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
              | | | |   || | | |||| |||||||||| ||||||||     ||    |  | | |
Db       1122 GACGAACCCGGCGACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGGAAGACCGT 1181

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
              |||||||||||||||||||||||||||||||| |||| | ||||||||||||||||||||
Db       1182 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCAGGAACTGCCGCGCCCCCCGCAA 1241

Qy        153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
              |||||||||||||   ||||||   || || ||||| ||||| || ||||| ||| | ||
Db       1242 GAAGGGCTGCTGGCGCTGCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACA 1301

Qy        213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
              ||| || || ||  | || || |||||||||| || || |||||| | || ||  |
Db       1302 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTC 1360

Qy        273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------A 326
              ||||||| | | ||||||||||||||||||  |||||  |||||  || | |       |
Db       1361 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAA 1420

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
              |||| ||| |  || || || |||||  | || || |       || | |||||||| ||
Db       1421 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCA 1480

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
              |||||| || ||||| || ||||| || |   | | | | ||||||||| ||||||||||
Db       1481 GATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGATCGGCGGCCAGCTCAAGGAGGC 1540
```





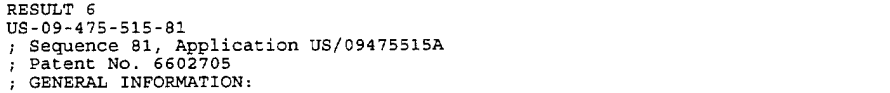
RESULT 5
US-09-475-515-80
; Sequence 80, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine

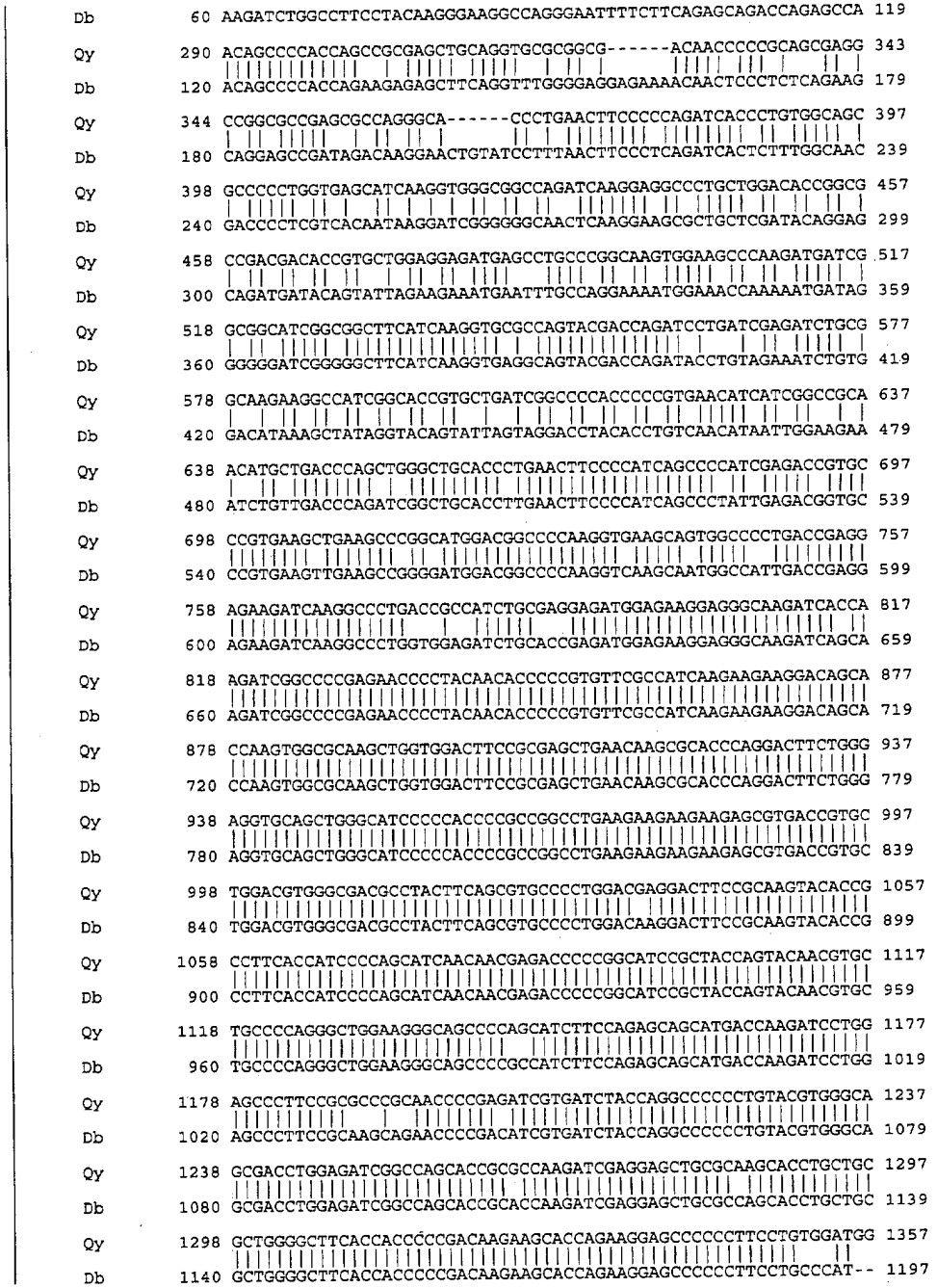
```
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
;   LENGTH: 2305
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(+).proinact.RTopt.YM
US-09-475-515-80

  Query Match             76.3%;  Score 1878.8;  DB 4;  Length 2305;
  Best Local Similarity   89.4%;  Pred. No. 2.5e-306;
  Matches 2061;  Conservative    0;  Mismatches  232;  Indels   13;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59
Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119
Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179
Qy        344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239
Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299
Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359
Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419
Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479
Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539
Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599
Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659
Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719
Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779
Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839
Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899
Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959
Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019
Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1079
Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1080 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1139
Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1140 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1199
Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1200 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1259
Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1260 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1319
Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1320 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1379
Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1380 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1439
Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1440 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1499
Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1500 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1559
Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1560 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1619
Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1620 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1679
Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1680 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1739
Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1740 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1799
Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1800 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1859
```

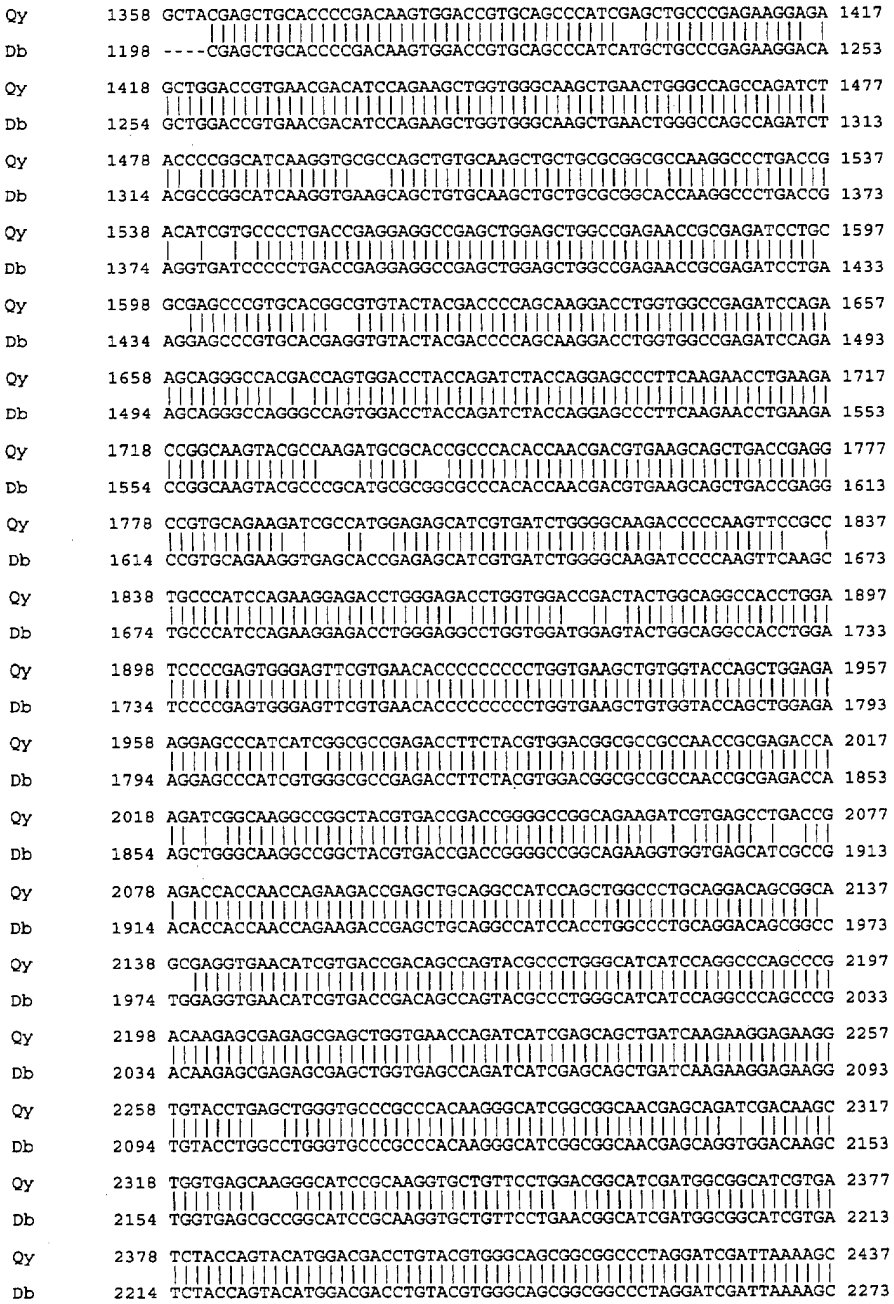


RESULT 6
US-09-475-515-81
; Sequence 81, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
; LENGTH: 2299
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).proinact.RTopt.YMWM
US-09-475-515-81

Query Match 75.2%; Score 1852; DB 4; Length 2299;
Best Local Similarity 89.0%; Pred. No. 7.6e-302;
Matches 2052; Conservative 0; Mismatches 235; Indels 19; Gaps 4;



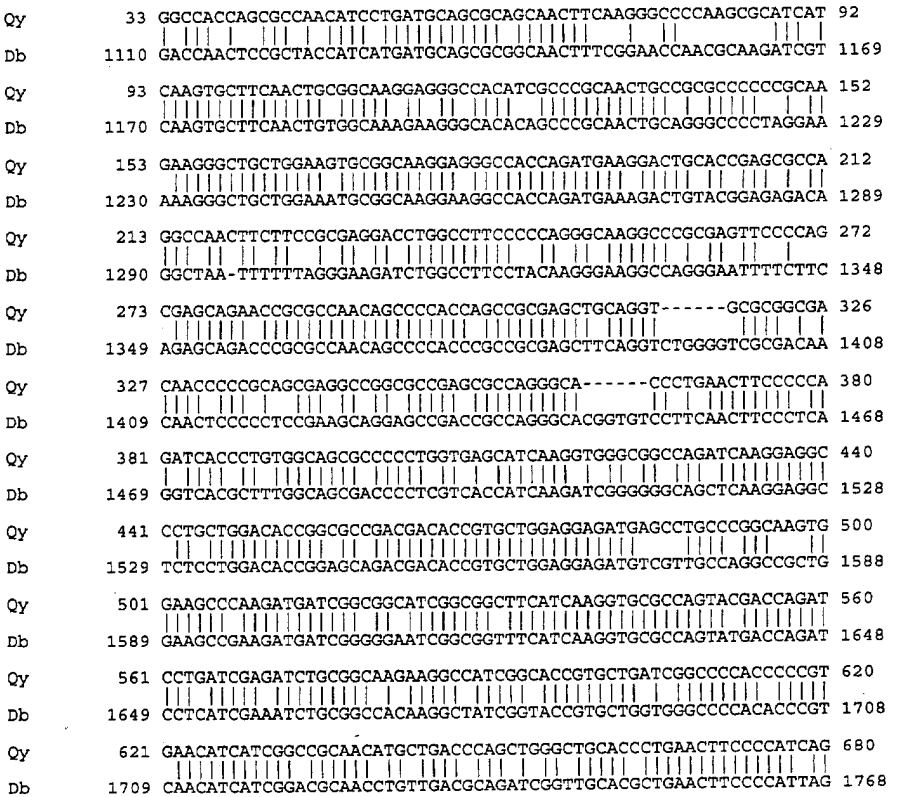


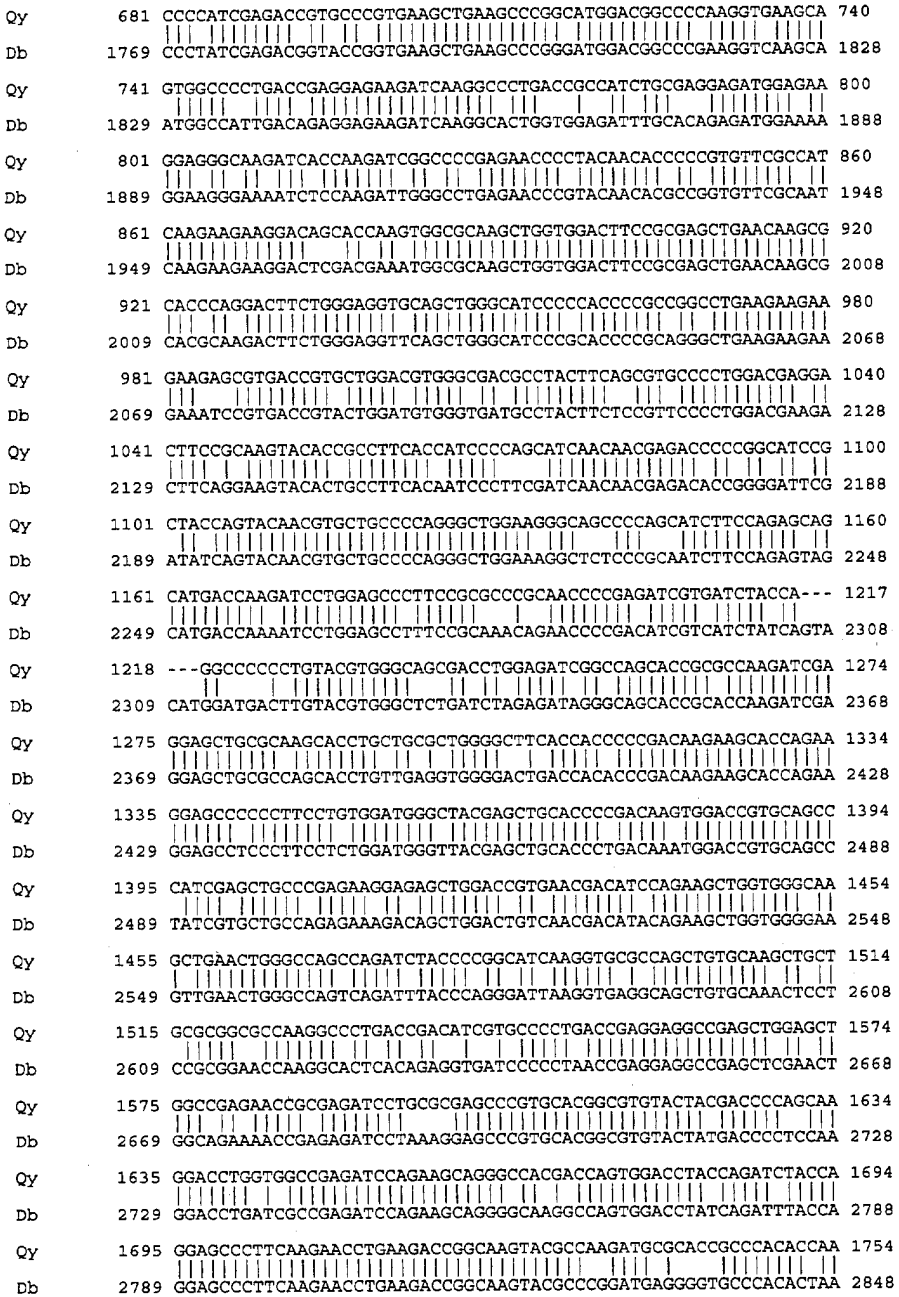


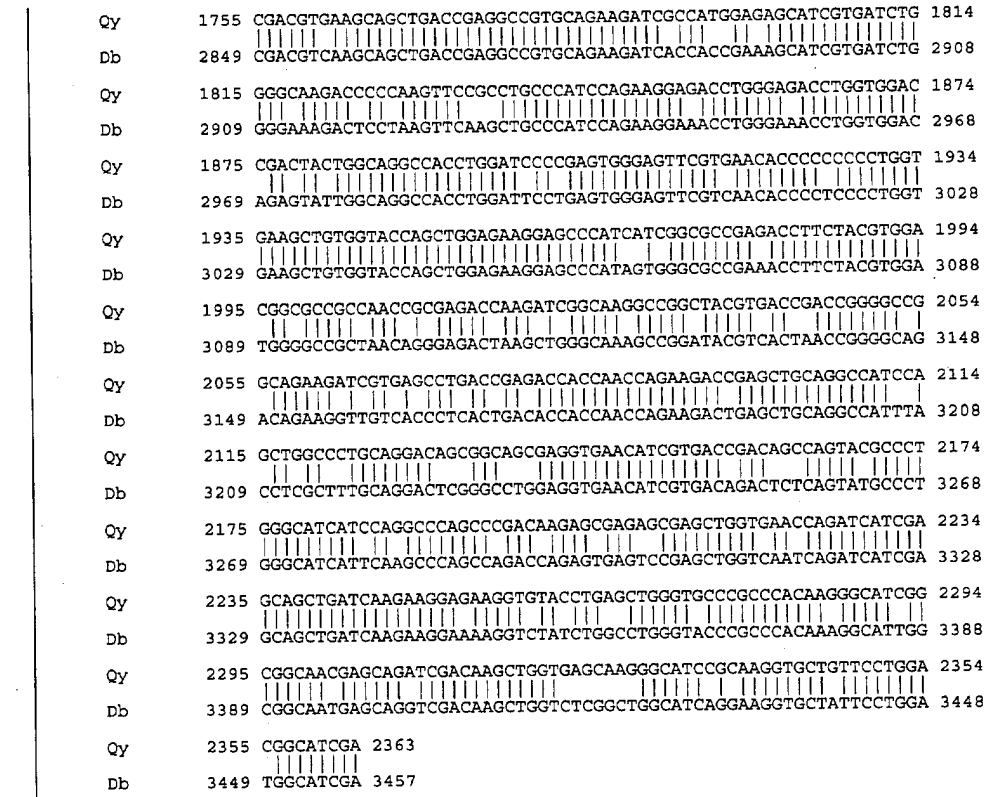
```
Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
             |||||||||||||||||||||||||||
Db       2274 TTCCCGGGGCTAGCACCGGTGAATTC 2299
```

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-SYNgp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 67.1%; Score 1651.8; DB 4; Length 4307;
Best Local Similarity 82.5%; Pred. No. 2.7e-268;
Matches 1939; Conservative 0; Mismatches 391; Indels 19; Gaps 4;

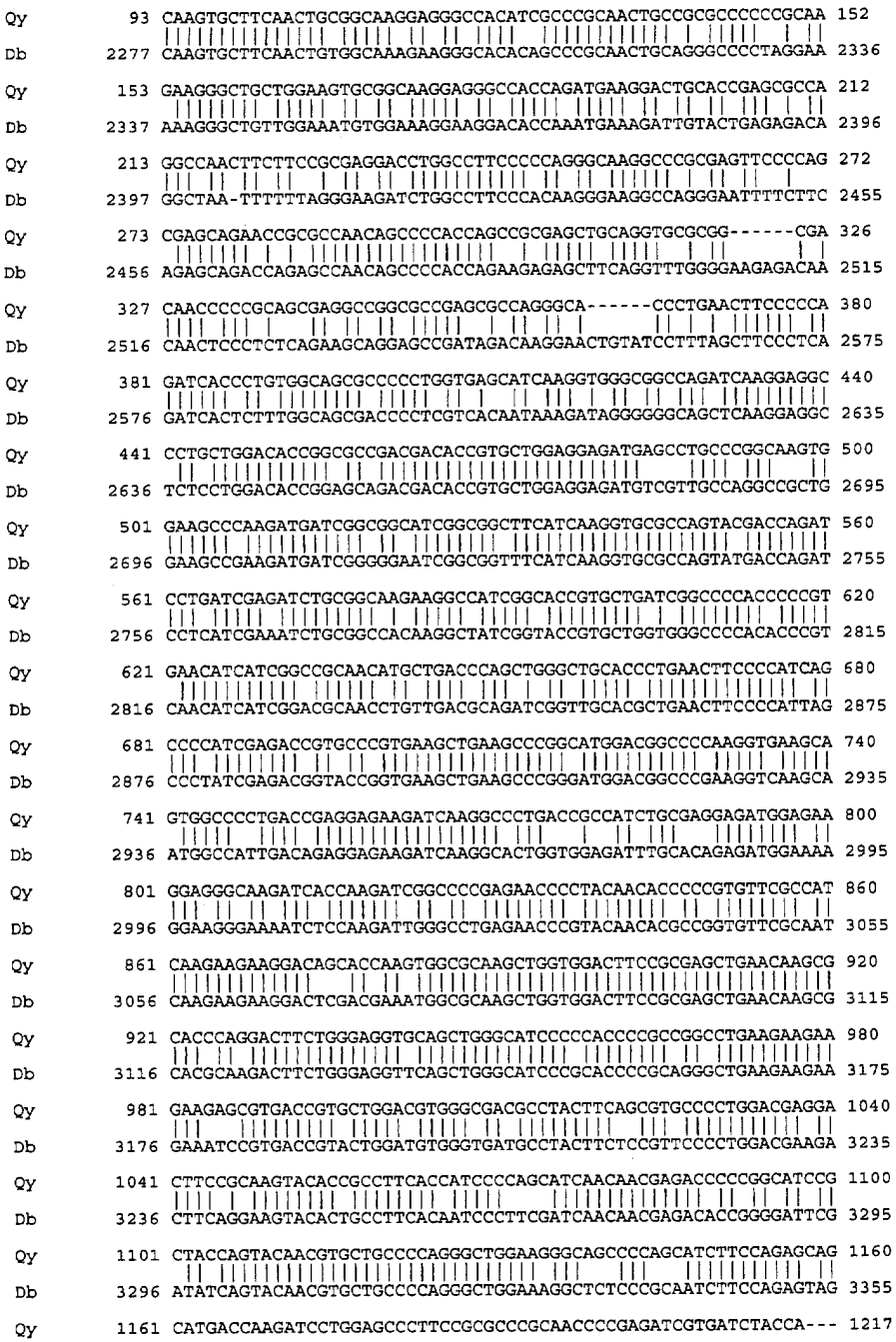






RESULT 8
US-09-552-950-5
; Sequence 5, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 9772
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:pSYNGP
US-09-552-950-5

Query Match 66.0%; Score 1624.6; DB 4; Length 9772;
Best Local Similarity 81.8%; Pred. No. 1.1e-263;
Matches 1922; Conservative 0; Mismatches 408; Indels 19; Gaps 4;


Qy 93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db 2277 CAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGCCCGCAACTGCAGGGCCCCTAGGAA 2336
Qy 153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db 2337 AAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACA 2396
Qy 213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db 2397 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 2455
Qy 273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db 2456 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAA 2515
Qy 327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db 2516 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCA 2575
Qy 381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db 2576 GATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGGGGGCAGCTCAAGGAGGC 2635
Qy 441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db 2636 TCTCCTGGACACCGGAGCAGACGACACCGTGCTGGAGGAGATGTCGTTGCCAGGCCGCTG 2695
Qy 501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db 2696 GAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAGAT 2755
Qy 561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db 2756 CCTCATCGAAATCTGCGGCCACAAGGCTATCGGTACCGTGCTGGTGGGCCCCACACCCGT 2815
Qy 621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db 2816 CAACATCATCGGACGCAACCTGTTGACGCAGATCGGTTGCACGCTGAACTTCCCCATTAG 2875
Qy 681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db 2876 CCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCGAAGGTCAAGCA 2935
Qy 741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db 2936 ATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACAGAGATGGAAAA 2995
Qy 801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db 2996 GGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCGGTGTTCGCAAT 3055
Qy 861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db 3056 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 3115
Qy 921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db 3116 CACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGGCTGAAGAAGAA 3175
Qy 981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db 3176 GAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCCCTGGACGAAGA 3235
Qy 1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db 3236 CTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACACCGGGGATTCG 3295
Qy 1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db 3296 ATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 3355
Qy 1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db 3356 CATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTCATCTATCAGTA 3415
Qy 1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db 3416 CATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCACCAAGATCGA 3475
Qy 1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db 3476 GGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAGAAGCACCAGAA 3535
Qy 1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db 3536 GGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGGACCGTGCAGCC 3595
Qy 1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db 3596 TATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAGCTGGTGGGGAA 3655
Qy 1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db 3656 GTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTGTGCAAACTCCT 3715
Qy 1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db 3716 CCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCCGAGCTCGAACT 3775
Qy 1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db 3776 GGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA 3835
Qy 1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db 3836 GGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTATCAGATTTACCA 3895
Qy 1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db 3896 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACACTAA 3955
Qy 1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db 3956 CGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGCATCGTGATCTG 4015
Qy 1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db 4016 GGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAAACCTGGTGGAC 4075
Qy 1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db 4076 AGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACCCCTCCCCTGGT 4135
Qy 1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db 4136 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA 4195
Qy 1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db 4196 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACTAACCGGGGCAG 4255
Qy 2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db 4256 ACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTGCAGGCCATTTA 4315
Qy 2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db 4316 CCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 4375
Qy 2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db 4376 GGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAATCAGATCATCGA 4435
Qy 2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294

```
Db       4436 GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCACAAAGGCATTGG 4495

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354

Db       4496 CGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 4555

Qy       2355 CGGCATCGA 2363

Db       4556 TGGCATCGA 4564


RESULT 9
US-09-872-733A-6
; Sequence 6, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733A
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 8366
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  DNA sequence
;   OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
;   OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
;   OTHER INFORMATION: resistance gene
US-09-872-733A-6

  Query Match             63.6%;  Score 1566;  DB 4;  Length 8366;
  Best Local Similarity   80.2%;  Pred. No. 6.7e-254;
  Matches 1885;  Conservative    0;  Mismatches  446;  Indels   19;  Gaps    4;

Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92

Db       1879 GACGAACTCGGCGACCATAATGATGCAGAGAGGCAACTTCCGGAACCAGCGGAAGATCGT 1938

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152

Db       1939 CAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCAGGAACTGCCGGGCCCCCCGGAA 1998

Qy        153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212

Db       1999 GAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACA 2058

Qy        213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272

Db       2059 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTC 2117

Qy        273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326

Db       2118 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAA 2177

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380

Db       2178 CAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCA 2237

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440

Db       2238 GATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGATCGGGGGGCAACTCAAGGAAGC 2297

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500

Db       2298 GCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTGCCAGGAAGATG 2357

Qy         501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560

Db       2358 GAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGAT 2417

Qy         561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620

Db       2418 ACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGT 2477

Qy         621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680

Db       2478 CAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAG 2537

Qy         681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740

Db       2538 CCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCA 2597

Qy         741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800

Db       2598 ATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCGAAATCTGTACAGAGATGGAGAA 2657

Qy         801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860

Db       2658 GGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCTACAACACTCCAGTCTTCGCAAT 2717

Qy         861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920

Db       2718 CAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAGCTGAACAAGAG 2777

Qy         921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980

Db       2778 AACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCCGCTGGGTTGAAGAAGAA 2837

Qy         981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040

Db       2838 GAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCCGTTCCCTTGGACGAGGA 2897

Qy        1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100

Db       2898 CTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAACGAGACACCAGGCATCCG 2957

Qy        1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160

Db       2958 CTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATCTTTCAAAGCAG 3017

Qy        1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217

Db       3018 CATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACCCAGACATCGTGATCTATCAGTA 3077

Qy        1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274

Db       3078 CATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGCACAGGACCAAGATCGA 3137

Qy        1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334

Db       3138 GGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACACCAGACAAGAAGCACCAGAA 3197

Qy        1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394

Db       3198 GGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATCCTGACAAGTGGACAGTGCAGCC 3257

Qy        1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454

Db       3258 CATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACGACATACAGAAGCTCGTGGGCAA 3317

Qy        1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514

Db       3318 GTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAGTTAGGCAGCTGTGCAAGCTGCT 3377

Qy        1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574

Db       3378 TCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAGGAAGCAGAGCTAGAACT 3437
```

```
Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db       3438 GGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATGGAGTGTACTACGACCCAAGCAA 3497

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db       3498 GGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCCAATGGACCTACCAAATCTACCA 3557

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db       3558 GGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGGTGCCCACACCAA 3617

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db       3618 CGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGCATCGTGATCTG 3677

Qy       1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db       3678 GGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACATGGGAGACATGGTGGAC 3737

Qy       1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCTGGT 1934
Db       3738 CGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTGAACACCCCTCCCTTGGT 3797

Qy       1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db       3798 GAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACCTTCTACGTGGA 3857

Qy       1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db       3858 TGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTACGTGACCAACCGAGGACG 3917

Qy       2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db       3918 ACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTGCAAGCCATCTA 3977

Qy       2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db       3978 CCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTGACAGACTCACAGTACGCACT 4037

Qy       2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db       4038 GGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAGAGCTGGTGAACCAGATCATCGA 4097

Qy       2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db       4098 GCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGGTACCAGCACACAAAGGAATTGG 4157

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db       4158 AGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGATCCGGAAGGTGCTGTTCCTGGA 4217

Qy       2355 CGGCATCGAT 2364
Db       4218 CGGGATCGAT 4227
```

RESULT 10
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

Query Match 62.1%; Score 1530.6; DB 4; Length 4338;
Best Local Similarity 80.1%; Pred. No. 5.2e-248;
Matches 1905; Conservative 0; Mismatches 444; Indels 28; Gaps 8;

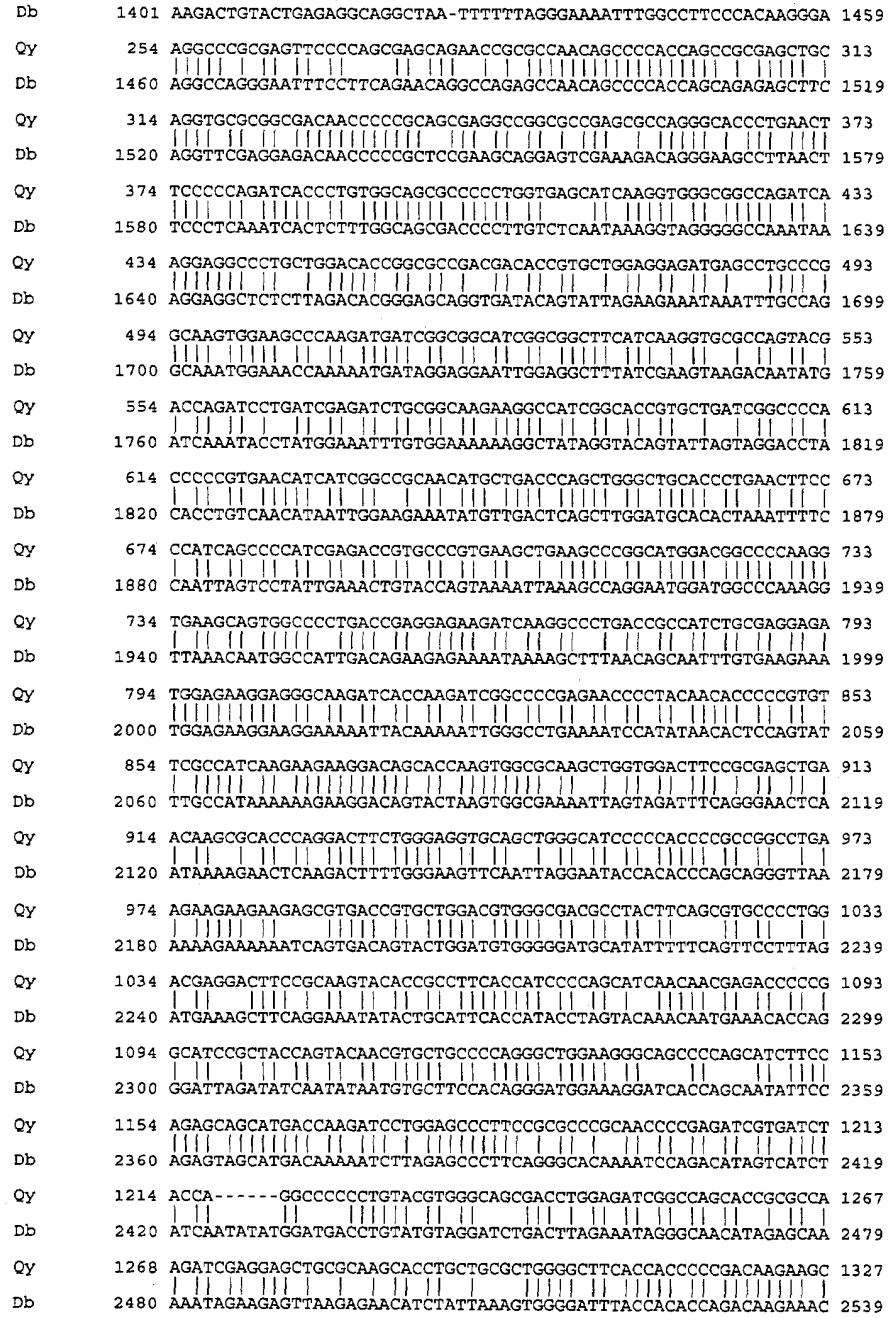
```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db       1085 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1144

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       1145 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1204

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       1205 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 1264

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       1265 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 1323

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       1324 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 1383

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       1384 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 1443

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       1444 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 1503

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db       1504 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 1563

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db       1564 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 1623

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db       1624 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683

Qy        599 TGCTGATCGGCCCCACC----CCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654
Db       1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATC 1743

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714
Db       1744 GGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCG 1803

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTG 774
Db       1804 GGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTA 1863

Qy        775 ACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAAC 834
Db       1864 GTCGAAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAAC 1923

Qy        835 CCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTG 894
```

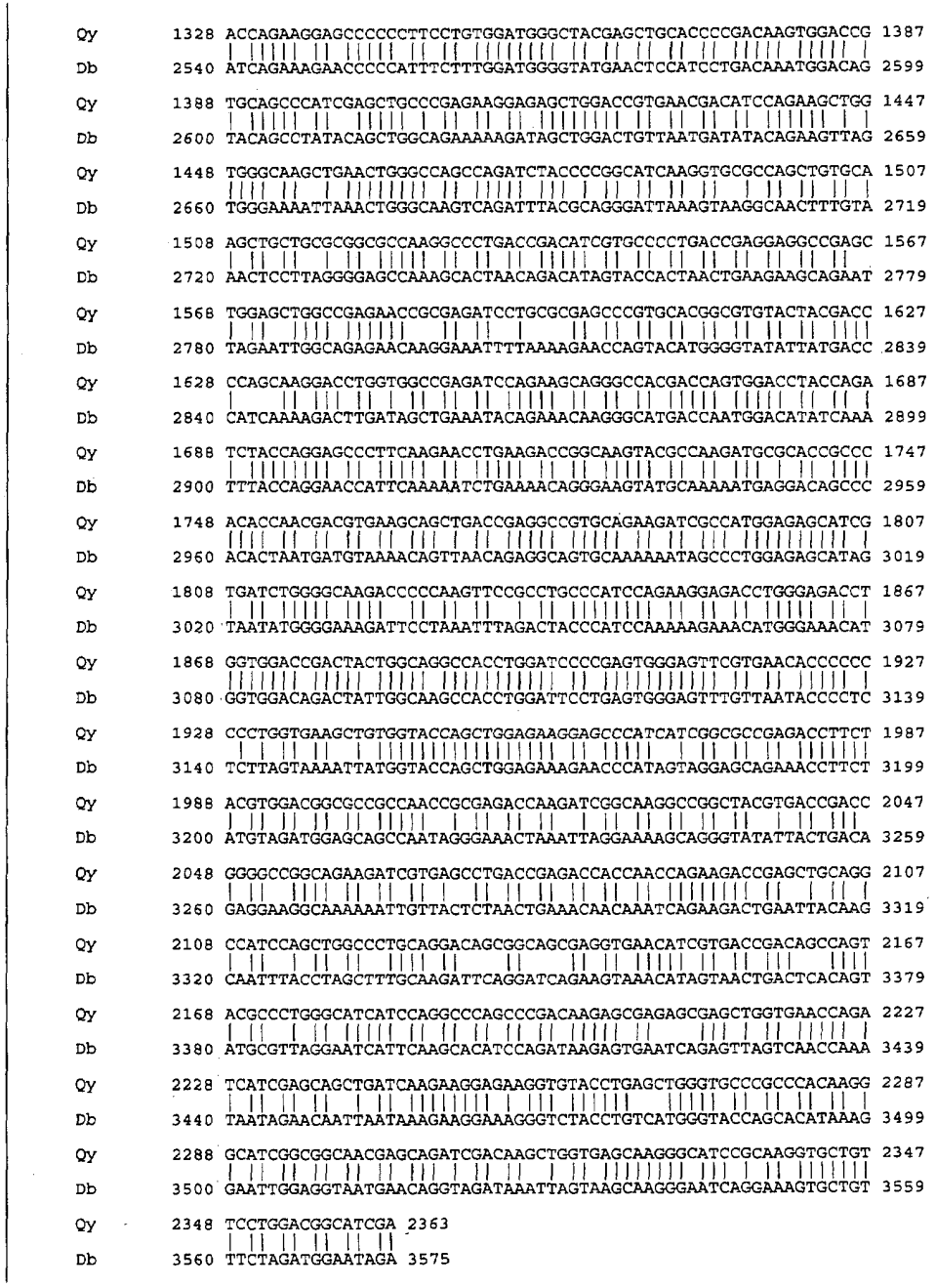
```
Db       1924 CCCTACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTG 1983
Qy        895 GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954
Db       1984 GTGGACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATC 2043
Qy        955 CCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC 1014
Db       2044 CCACATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCC 2103
Qy       1015 TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGC 1074
Db       2104 TACTTCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGC 2163
Qy       1075 ATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAG 1134
Db       2164 ATCAACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAG 2223
Qy       1135 GGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGC 1194
Db       2224 GGATCACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAA 2283
Qy       1195 AACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAG 1248
Db       2284 AACCCAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAG 2343
Qy       1249 ATC-GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTT 1307
Db       2344 ATCGGGGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACT 2403
Qy       1308 CACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCT 1367
Db       2404 GACCACACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACT 2463
Qy       1368 GCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGT 1427
Db       2464 GCATCCTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGT 2523
Qy       1428 GAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCAT 1487
Db       2524 GAACGACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCAT 2583
Qy       1488 CAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCC 1547
Db       2584 CAAAGTTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCC 2643
Qy       1548 CCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGT 1607
Db       2644 ACTGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGT 2703
Qy       1608 GCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCA 1667
Db       2704 ACATGGAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCA 2763
Qy       1668 CGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTA 1727
Db       2764 AGGCCAATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTA 2823
Qy       1728 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1787
Db       2824 CGCAAGGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAA 2883
Qy       1788 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1847
Db       2884 GATCACCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943
Qy       1848 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1907
Db       2944 GAAGGAGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTG 3003
Qy       1908 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1967
Db       3004 GGAGTTCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCAT 3063

Qy       1968 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2027
Db       3064 CGTGGGAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAA 3123
Qy       2028 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2087
Db       3124 GGCAGGCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAA 3183
Qy       2088 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2147
Db       3184 CCAGAAGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAA 3243
Qy       2148 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2207
Db       3244 CATCGTGACAGACTCACAGTACG-CATGGGCATCATCCAAGCACAACCAGACCAATCCGA 3302
Qy       2208 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2267
Db       3303 GTCAGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGC 3362
Qy       2268 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2327
Db       3363 ATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGC 3422
Qy       2328 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db       3423 TGGGATCCGGAAGGTGCTGTTCCTGGACGGGATCGAT 3459
```

RESULT 11
US-09-184-418C-8
; Sequence 8, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 8
; LENGTH: 9010
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=96ZM651; 137..1621:"gag"; 1426..4425:"pol";
; OTHER INFORMATION: 4370..4948:"vif"; 4888..5178:"vpr";
; OTHER INFORMATION: 5159..5373-7734..7824:"tat"; 5298..5373-7734..7981:"rev";
; OTHER INFORMATION: 5387..5647:"vpu"; 5565..8171:"env"; 8173..8793:"nef"
US-09-184-418C-8

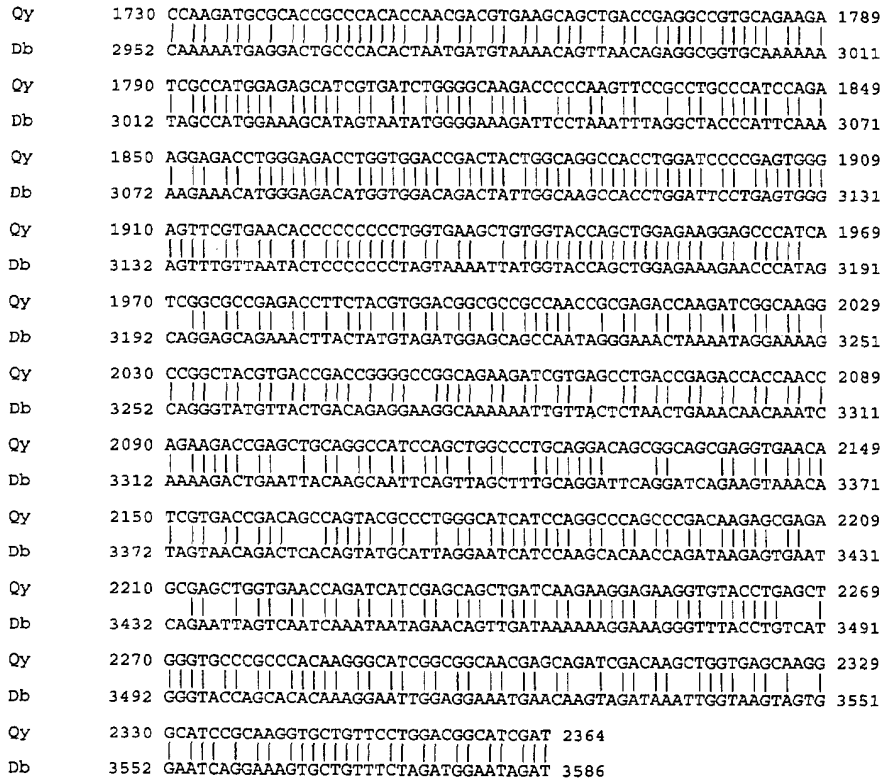
Query Match 48.9%; Score 1203.6; DB 4; Length 9010;
Best Local Similarity 70.0%; Pred. No. 3.2e-193;
Matches 1650; Conservative 0; Mismatches 699; Indels 7; Gaps 2;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1221 TGGCTGAGGCAATGAGCCAAACAAATAGTGTAAACATACTGATGCAGAAAAGCAATTTTA 1280
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1281 AAGGAAATAAAAGAATGGTTAAATGTTTTAACTGTGGTAAGGAAGGGCACATAGCCAGAA 1340
Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1341 ATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAGGGACACCAAATGA 1400
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
```



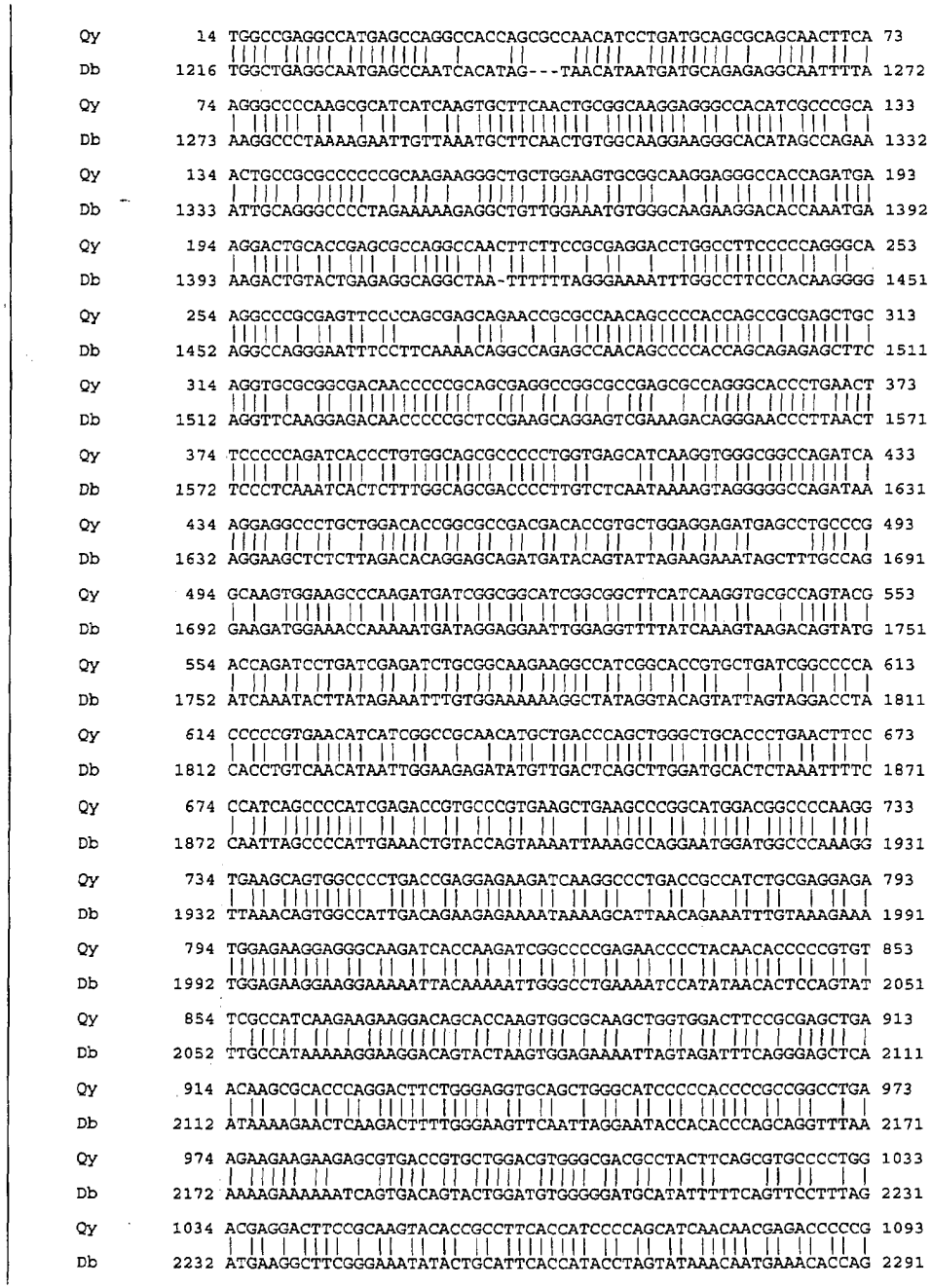


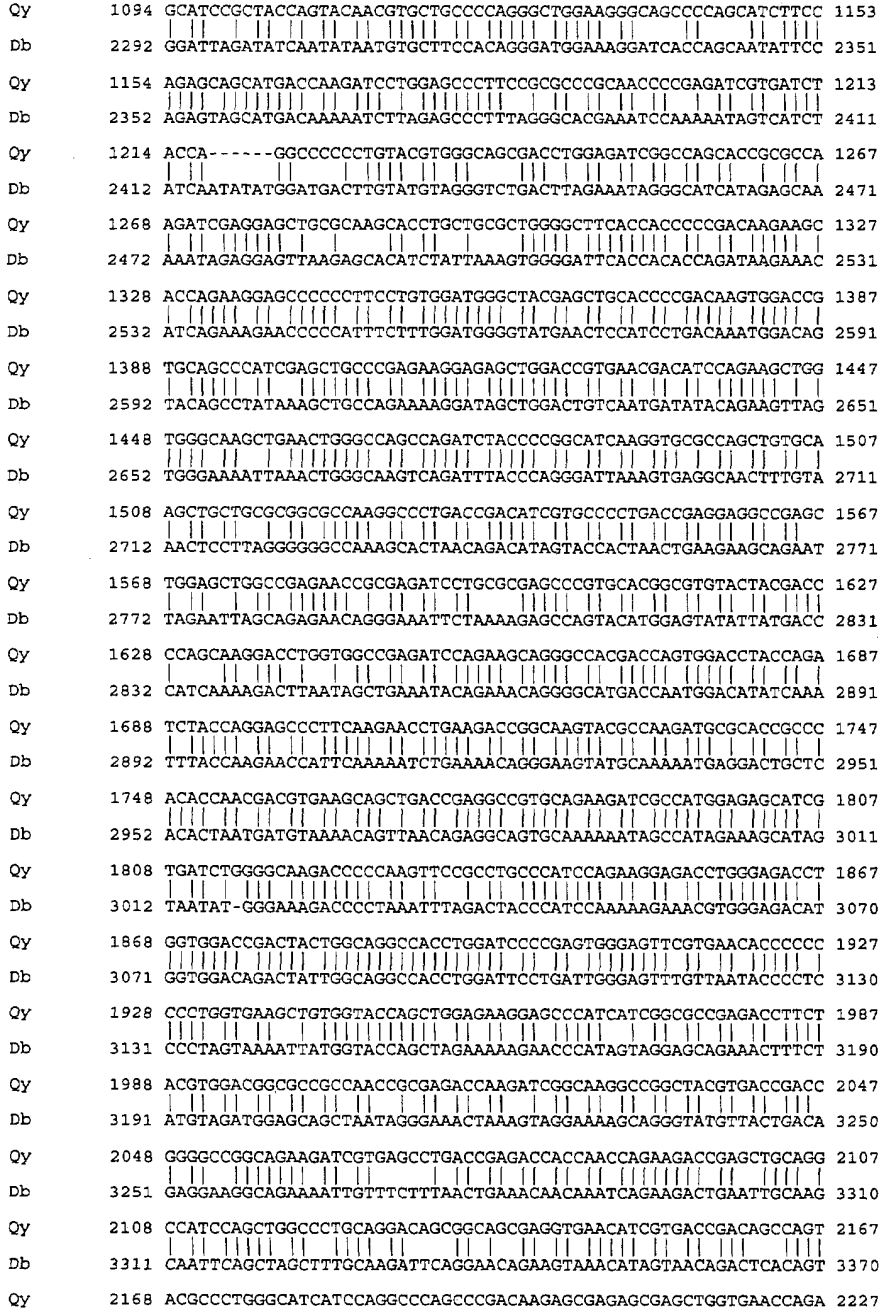
```
RESULT 12
US-09-184-418C-9
; Sequence 9, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
;  APPLICANT: Hahn, Beatrice
;  APPLICANT:  Gao, Feng
;  APPLICANT:  Shaw, George
;  TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
;  TITLE OF INVENTION:  IMMUNODEFICIENCY VIRUS TYPE 1
;  FILE REFERENCE: D6287
;  CURRENT APPLICATION NUMBER: US/09/184,418C
;  CURRENT FILING DATE:  1999-11-02
;  NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 9
;   LENGTH: 8972
;   TYPE: DNA
;   ORGANISM: Human immunodeficiency virus type 1
;   FEATURE:
;   OTHER INFORMATION: isolate=96ZM751.3; 137.1632:gag; 1419.4435:pol;
;   OTHER INFORMATION: 4380.4958:vif; 4898.5188:vpr; 5169.7814:tat;
;   OTHER INFORMATION: 5308.7938:rev; 5407.5667:vpu; 5585.8128:env;
;   OTHER INFORMATION: 8130.8753:nef
US-09-184-418C-9

  Query Match             47.6%;  Score 1172.6;  DB 4;  Length 8972;
  Best Local Similarity   69.7%;  Pred. No. 4.9e-188;
  Matches 1655;  Conservative    0;  Mismatches  694;  Indels   26;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1214 TGGCTGAAGCAATGAGCCAAGTAAACAATACAAACATAATGATGCAGAAAAGCAATTTTA 1273

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1274 AAGGCCCTAAAAGAATTGTTAAATGTTTCAACTGTGGCAGGGAAGGGCATATAGCCAGGA 1333

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1334 ATTGCAGGGCTCCTGGGAAAAAAGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGA 1393

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       1394 AAGACTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAAATTTGGCCTTCCCAGAAGGGG 1452

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAG----------------AACCGCGCCAACAGCC 295
Db       1453 AGGCCGGGGAACTTCCTTCAGAACAGACCAGAGCCAACAGCCCCACCAGCTCCAACAGCC 1512

Qy        296 CCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGC 355
Db       1513 CCACCAGCAGAGAGCTTCAGGTTCGAGGAGACAACCCCTGCCCCGAGGCAGGAGCAGAAA 1572

Qy        356 GCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCA 415
Db       1573 GACAAGGAACCCTTAACTGCCCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAA 1632

Qy        416 AGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGG 475
Db       1633 AAGTAGGGGGTCAGATAAAGGAGGCTCTCTTGGATACAGGAGCAGATGATACAGTATTAG 1692

Qy        476 AGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA 535
Db       1693 AAGAAATAAATTTGCCAGGAAAATGGAAACCAAAAATGATAGGAGGAATTGGAGGTTTTA 1752

Qy        536 TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCA 595
Db       1753 TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCTATAGGTA 1812

Qy        596 CCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGG 655
Db       1813 CAGTATTAGTAGGACCTACACCTGTCAACATAATTGGGAGAAATATGTTGACCCAGCTTG 1872

Qy        656 GCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCG 715
Db       1873 GCTGCACACTAAATTTTCCAATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAG 1932

Qy        716 GCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA 775
Db       1933 GAATGGATGGCCCAAGGGTCAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAA 1992

Qy        776 CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACC 835
Db       1993 CAGCAATTTGTGAAGAAATGGAAAAGGAAGGAAAAATTACAAAAATTGGCCCTGAGAATC 2052

Qy        836 CCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGG 895
Db       2053 CATATAACACTCCAGTATTTGCCATAAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAG 2112

Qy        896 TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCC 955
Db       2113 TAGATTTCAGGGAACTCAATAAAAGAACTCAGGACTTTTGGGAAGTTCAATTAGGAATAC 2172

Qy        956 CCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT 1015
Db       2173 CACACCCAGCGGGGTTAAAAAAGAAAAAGTCAGTGACAGTACTGGATGTGGGGGATGCGT 2232

Qy       1016 ACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCA 1075
Db       2233 ATTTTTCAGTTCCTTTAGATGAAGGCTTCAGGAAATATACTGCATTCACCATACCTAGTA 2292

Qy       1076 TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGG 1135
Db       2293 TAAACAATGAAACACCTGGGATTAGATATCAATATAATGTGCTTCCACAGGGATGGAAAG 2352

Qy       1136 GCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCA 1195
Db       2353 GATCACCATCAATATTCCAGAGTAGCATGATAAAAATCTTAGAGCCCTTTAGGACACAAA 2412

Qy       1196 ACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGA 1249
Db       2413 ACCCAGAAATAGTTATCTATCAATATATGGATGACTTGTATGTAGGATCTGATTTAGAAA 2472

Qy       1250 TCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCA 1309
Db       2473 TAGGGCAACACAGAGCAAAAATAGAGGAGTTAAGAGAACACCTATTGAGATGGGGATTTA 2532

Qy       1310 CCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGC 1369
Db       2533 CTACACCAGACAAGAAGCATCAGAAAGAGCCCCCATTTCTTTGGATGGGGTATGAACTCC 2592

Qy       1370 ACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGA 1429
Db       2593 ATCCTGACAAATGGACAGTACAGCCTATAAAGCTGCCAGAAAAGGAGAGCTGGACTGTCA 2652

Qy       1430 ACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCA 1489
Db       2653 ATGATATACAGAAGTTAGTGGGAAAATTAAACT-GGCAAGTCAGATTTACGCAGGGATTA 2711

Qy       1490 AGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCC 1549
Db       2712 AAGTAAGGCAACTGTGTAAACTCCTTAGGGGAGCCAAAGCACTAACAGACATAGTACCAT 2771

Qy       1550 TGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGC 1609
Db       2772 TGACTGAAGAGGCAGAATTAGAATTGGCAGAGAGCAGGGAAATTCTAAAAGAACCAGTAC 2831

Qy       1610 ACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACG 1669
Db       2832 ATGGAGTATATTATGACCCATCAAAAGACTTAATAGCTGAAATACAGAAACAAGGGCATG 2891

Qy       1670 ACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACG 1729
Db       2892 ACCAATGGACATATCAAGTTTACCAAGAACCATTCAAAAATCTGAAAACAGGAAAGTATG 2951
```

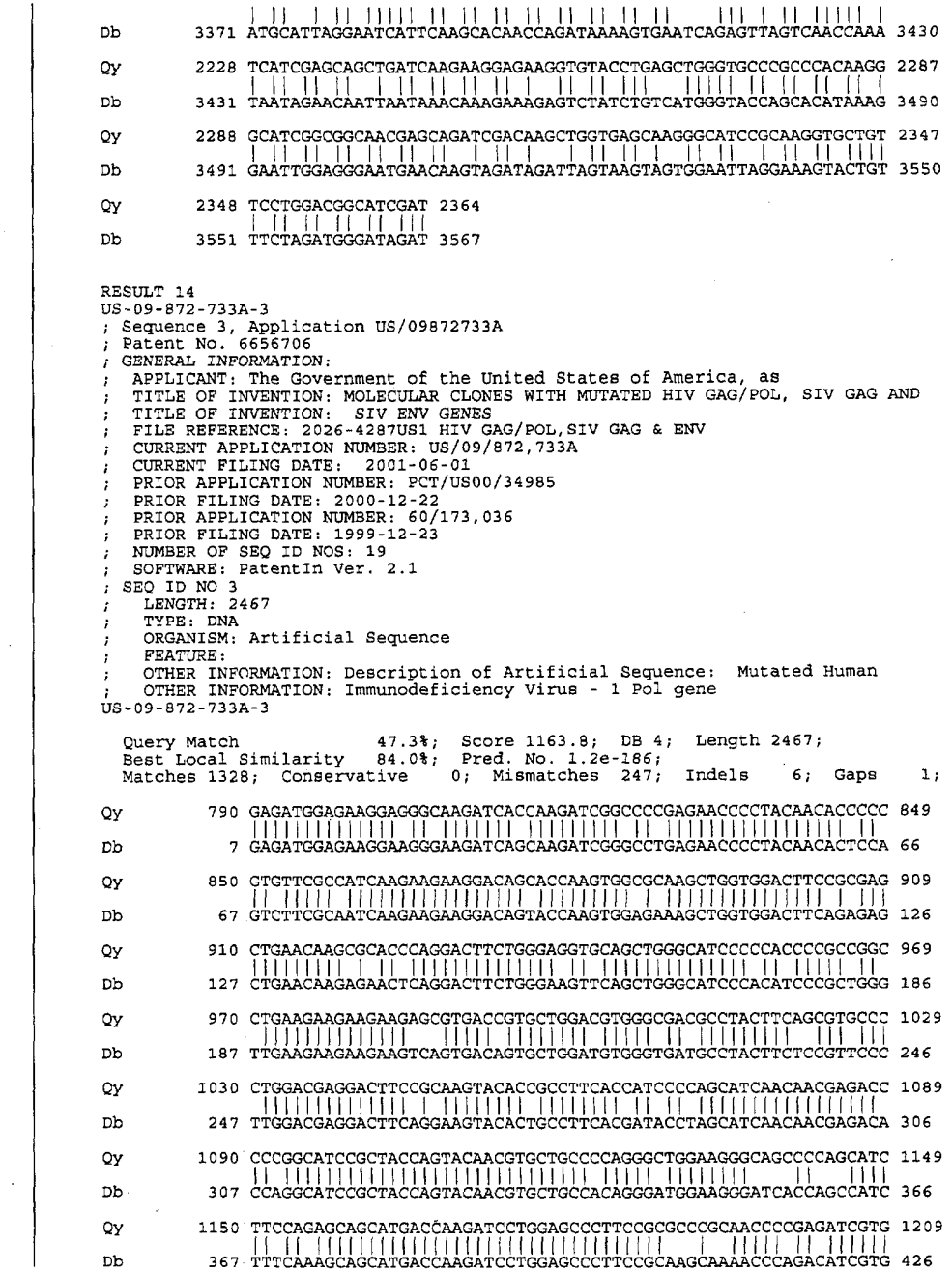


RESULT 13
US-09-184-418C-11
; Sequence 11, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 11
; LENGTH: 8959
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=94IN476.104; 138.1613:"gag";
; OTHER INFORMATION: 1418.4428:"pol"; 4361.4939:"vif"; 4879.5169:"vpr";
; OTHER INFORMATION: 5150.7782:tat"; 5289.7939:"rev"; 5378.5638:"vpu";
; OTHER INFORMATION: 5556.8129:"env"; 8131.8754:nef
US-09-184-418C-11

Query Match 47.3%; Score 1165.4; DB 4; Length 8959;
Best Local Similarity 69.6%; Pred. No. 7.9e-187;
Matches 1640; Conservative 0; Mismatches 706; Indels 11; Gaps 4;







RESULT 14
US-09-872-733A-3
; Sequence 3, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Pol gene
US-09-872-733A-3

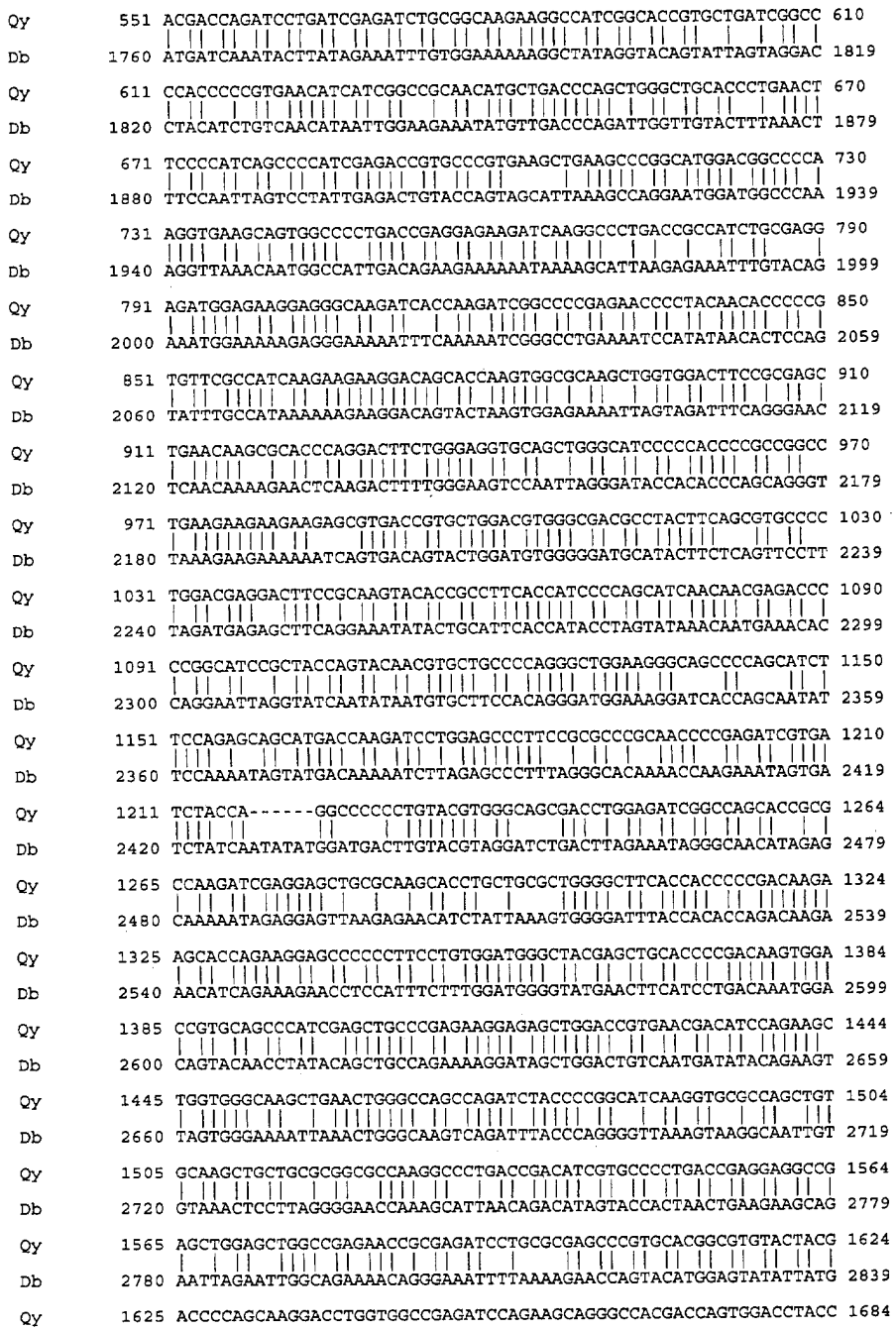
Query Match 47.3%; Score 1163.8; DB 4; Length 2467;
Best Local Similarity 84.0%; Pred. No. 1.2e-186;
Matches 1328; Conservative 0; Mismatches 247; Indels 6; Gaps 1;

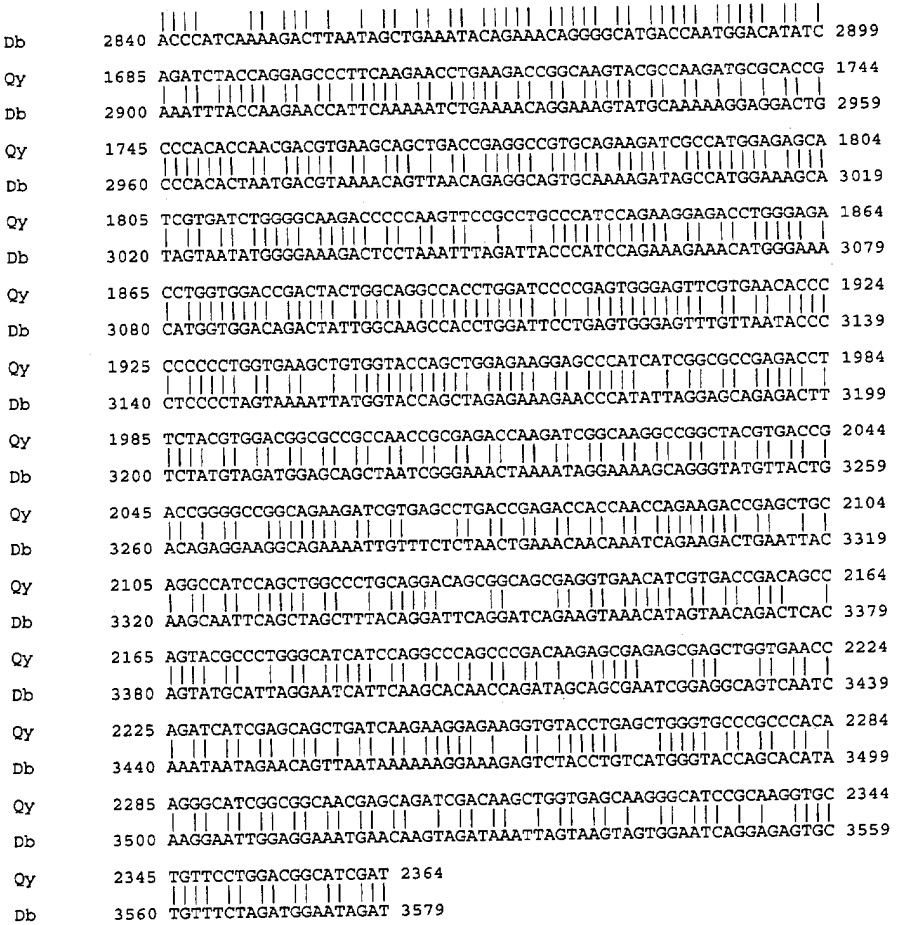
```
Qy       1210 ATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGC 1263
Db        427 ATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGCACAGG 486

Qy       1264 GCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAG 1323
Db        487 ACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACACCAGACAAG 546

Qy       1324 AAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGG 1383
Db        547 AAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATCCTGACAAGTGG 606

Qy       1384 ACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAG 1443
Db        607 ACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACGACATACAGAAG 666

Qy       1444 CTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTG 1503
Db        667 CTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAGTTAGGCAGCTG 726

Qy       1504 TGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCC 1563
Db        727 TGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAGGAAGCA 786

Qy       1564 GAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTAC 1623
Db        787 GAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATGGAGTGTACTAC 846

Qy       1624 GACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTAC 1683
Db        847 GACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCCAATGGACCTAC 906

Qy       1684 CAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACC 1743
Db        907 CAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGGT 966

Qy       1744 GCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGC 1803
Db        967 GCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGC 1026

Qy       1804 ATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAG 1863
Db        1027 ATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACATGGGAG 1086

Qy       1864 ACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACC 1923
Db        1087 ACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTGAACACC 1146

Qy       1924 CCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACC 1983
Db        1147 CCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACC 1206

Qy       1984 TTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACC 2043
Db        1207 TTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTACGTGACC 1266

Qy       2044 GACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTG 2103
Db        1267 AACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTG 1326

Qy       2104 CAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGC 2163
Db        1327 CAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTGACAGACTCA 1386

Qy       2164 CAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAAC 2223
Db        1387 CAGTACGCACTGGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAGAGCTGGTGAAC 1446

Qy       2224 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCAC 2283
Db        1447 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGGTACCAGCACAC 1506

Qy       2284 AAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2343
Db        1507 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGATCCGGAAGGTG 1566

Qy       2344 CTGTTCCTGGACGGCATCGAT 2364
Db        1567 CTGTTCCTGGACGGGATCGAT 1587
```

RESULT 15
US-09-184-418C-4
; Sequence 4, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 4
; LENGTH: 8992
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=92RW009; 139.1624:gag; 1690.4428:pol(N-terminus uncertai
; OTHER INFORMATION: 4373.4951:vif; 4891.5181:vpr; 5162.7801:tat; 5301.7958:rev;
; OTHER INFORMATION: 5403.5648:vpu; 5566.8148:env; 8150.8773:nef
US-09-184-418C-4

Query Match 46.4%; Score 1142; DB 4; Length 8992;
Best Local Similarity 68.6%; Pred. No. 6.5e-183;
Matches 1620; Conservative 0; Mismatches 730; Indels 10; Gaps 3;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1221 TGGCTGAAGCAATGAGCCAAGTACAACAACCAAACATAATGATGCAGAGAGGCAATTTTA 1280

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1281 AGGGCCAGAGAAGAATTATTAAGTGTTTCAACTGTGGCAAAGAAGGACACCTAGCCAGAA 1340

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1341 ATTGCAGGGCCCCTAGAAAAAAGGGCTGTTGGAAATGTGGAAAGGAGGGACACCAAATGA 1400

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       1401 AAGACTGCACTGAGAGACAGGCTAA-TTTTTTAGGGAAAATTTGGCCTTCCAACAAGGGG 1459

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCT-- 311
Db       1460 AGGCCAGGAAATTTTCCCCAGAGCAGACTGGAGCCAACAGCCCCACCAGCAGAGAACTTT 1519

Qy        312 -GCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGA 370
Db       1520 GGAATGGGGGAAGAGATAGCCTCTCCTCTGAAACAGGAGCAGAAAGACAGGGAACCTTTA 1579

Qy        371 ACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGA 430
Db       1580 ATTTCCCTCAAATCACTCTTTGGCAACGACCCCTTGTCACAGTAAAAATAGGAGGTCAGC 1639

Qy        431 TCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGC 490
Db       1640 TAAGAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATAAATTTGC 1699

Qy        491 CCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGT 550
Db       1700 CAGGAAAATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAGGTAAAACAGT 1759
```





Search completed: April 10, 2004, 16:22:46
Job time : 127.333 secs

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 12:49:38 ; Search time 593.333 Seconds
(without alignments)
15572.780 Million cell updates/sec

Title: US-09-610-313-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2475585 seqs, 1875730760 residues

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2455.4	99.7	2457	14	US-10-190-435-44	Sequence 44, Appl
2	2455.4	99.7	2457	14	US-10-190-305A-38	Sequence 38, Appl
3	2430.2	98.7	2457	14	US-10-190-435-45	Sequence 45, Appl
4	2430.2	98.7	2457	14	US-10-190-305A-39	Sequence 39, Appl
5	2422.6	98.4	2445	14	US-10-190-435-43	Sequence 43, Appl
6	2422.6	98.4	2445	14	US-10-190-305A-37	Sequence 37, Appl
7	2415.6	98.1	3930	14	US-10-190-435-9	Sequence 9, Appli
8	2414	98.0	3930	14	US-10-190-435-10	Sequence 10, Appl
9	2414	98.0	3930	14	US-10-190-435-11	Sequence 11, Appl
10	2414	98.0	5184	14	US-10-190-435-58	Sequence 58, Appl
11	2414	98.0	5184	14	US-10-190-305A-82	Sequence 82, Appl
12	2383.6	96.8	3531	14	US-10-190-435-13	Sequence 13, Appl
13	2382	96.7	3537	14	US-10-190-435-14	Sequence 14, Appl
14	2382	96.7	3537	14	US-10-190-435-15	Sequence 15, Appl
15	2381	96.7	5145	14	US-10-190-435-12	Sequence 12, Appl
16	2381	96.7	5145	14	US-10-190-305A-12	Sequence 12, Appl
17	2322.6	94.3	3607	14	US-10-190-435-48	Sequence 48, Appl
18	2322.6	94.3	3607	14	US-10-190-305A-42	Sequence 42, Appl
19	2322.6	94.3	3624	14	US-10-190-435-47	Sequence 47, Appl
20	2322.6	94.3	3624	14	US-10-190-305A-41	Sequence 41, Appl
21	2304.4	93.6	3597	14	US-10-190-435-46	Sequence 46, Appl
22	2304.4	93.6	3597	14	US-10-190-305A-40	Sequence 40, Appl
23	2142	87.0	2466	14	US-10-241-009-31	Sequence 31, Appl
24	2142	87.0	2466	14	US-10-190-434B-31	Sequence 31, Appl
25	2142	87.0	2466	14	US-10-190-305A-33	Sequence 33, Appl
26	2122.8	86.2	2472	14	US-10-241-009-32	Sequence 32, Appl
27	2122.8	86.2	2472	14	US-10-190-434B-32	Sequence 32, Appl
28	2122.8	86.2	2472	14	US-10-190-305A-34	Sequence 34, Appl
29	2115.2	85.9	2460	14	US-10-241-009-30	Sequence 30, Appl
30	2115.2	85.9	2460	14	US-10-190-434B-30	Sequence 30, Appl
31	2115.2	85.9	2460	14	US-10-190-305A-32	Sequence 32, Appl
32	2114.2	85.8	3564	14	US-10-241-009-13	Sequence 13, Appl
33	2114.2	85.8	3564	14	US-10-241-009-14	Sequence 14, Appl
34	2114.2	85.8	3564	14	US-10-190-434B-13	Sequence 13, Appl
35	2114.2	85.8	3564	14	US-10-190-434B-14	Sequence 14, Appl
36	2113.6	85.8	4716	14	US-10-190-435-17	Sequence 17, Appl
37	2113.6	85.8	4716	14	US-10-190-305A-13	Sequence 13, Appl
38	2109.8	85.7	3999	14	US-10-241-009-9	Sequence 9, Appli
39	2109.8	85.7	3999	14	US-10-190-434B-9	Sequence 9, Appli
40	2108.2	85.6	3999	14	US-10-241-009-10	Sequence 10, Appl
41	2108.2	85.6	3999	14	US-10-241-009-11	Sequence 11, Appl
42	2108.2	85.6	3999	14	US-10-190-434B-10	Sequence 10, Appl
43	2108.2	85.6	3999	14	US-10-190-434B-11	Sequence 11, Appl
44	2108.2	85.6	5283	14	US-10-241-009-54	Sequence 54, Appl
45	2108.2	85.6	5283	14	US-10-190-434B-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44

Query Match 99.7%; Score 2455.4; DB 14; Length 2457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
```

```
Db       121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Qy       181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db       181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
```

```
Db      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
```

```
Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 2
US-10-190-305A-38
; Sequence 38, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-305A-38

  Query Match             99.7%;  Score 2455.4;  DB 14;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2456;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
```

```
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 3
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 45
;  LENGTH: 2457
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

  Query Match            98.7%;  Score 2430.2;  DB 14;  Length 2457;
  Best Local Similarity  99.6%;  Pred. No. 0;
  Matches 2448;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
```

```
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 4
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

Query Match 98.7%; Score 2430.2; DB 14; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60
Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120
Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180
Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240
```

```
Qy       247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db       241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy       307 GAGCTGCAGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db       301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy       367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db       361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy       427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db       421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy       487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db       481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy       547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db       541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy       607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db       601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy       667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db       661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy       727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db       721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy       787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db       781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy       847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db       841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy       907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db       901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy       967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db       961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy      1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db      1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy      1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db      1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy      1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db      1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy      1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db      1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy      2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 5
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMWM_C
US-10-190-435-43

  Query Match             98.4%;  Score 2422.6;  DB 14;  Length 2445;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2441;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              |||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334

Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394

Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445


RESULT 6
US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: p2Pol.opt.YMWM_C
US-10-190-305A-37

  Query Match            98.4%;  Score 2422.6;  DB 14;  Length 2445;
  Best Local Similarity  99.6%;  Pred. No. 0;
  Matches 2441;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

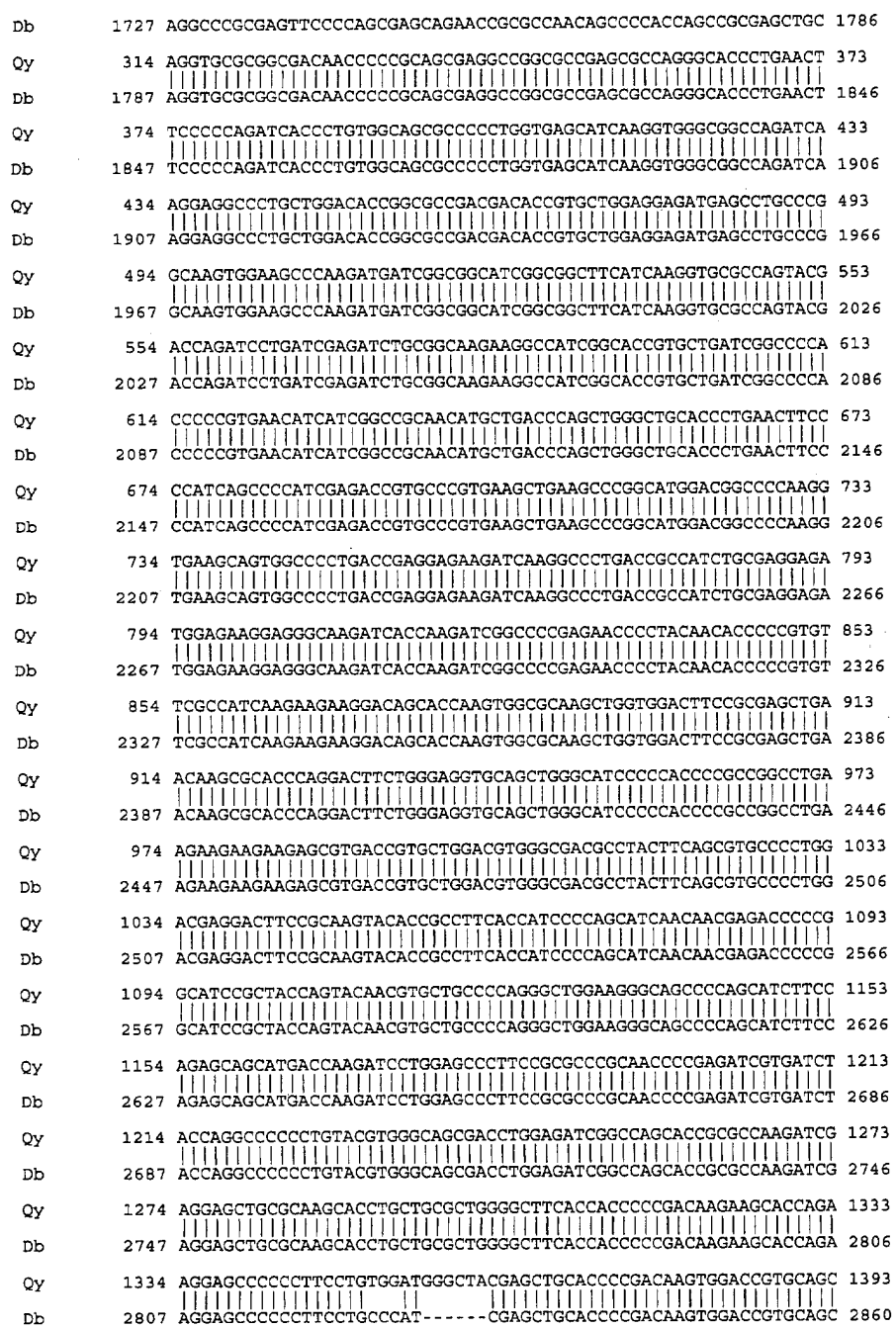
```
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260
Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320
Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374
Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434
Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494
Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554
Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614
Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674
Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734
Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794
Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854
Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914
Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974
Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034
Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094
```

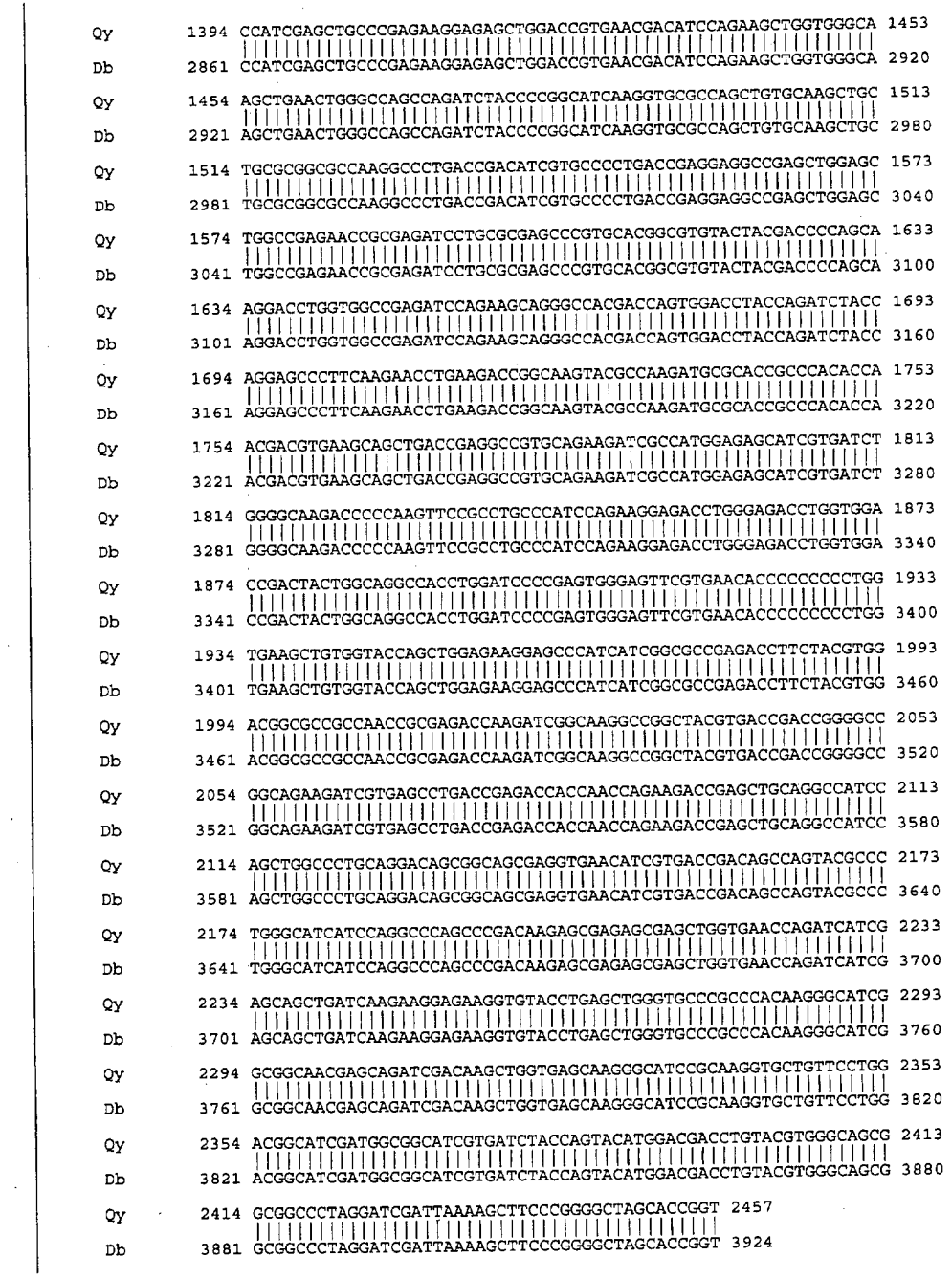
```
Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154
Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214
Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274
Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334
Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394
Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 7
US-10-190-435-9
; Sequence 9, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagCompIPolmut_C
US-10-190-435-9

Query Match 98.1%; Score 2415.6; DB 14; Length 3930;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2434; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606
Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
```


Db 1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db 2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746
Qy 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db 2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806
Qy 1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db 2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860


Qy 1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
Db 2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920
Qy 1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db 2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980
Qy 1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db 2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040
Qy 1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db 3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100
Qy 1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db 3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160
Qy 1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db 3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220
Qy 1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db 3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280
Qy 1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
Db 3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340
Qy 1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db 3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400
Qy 1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db 3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460
Qy 1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db 3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520
Qy 2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db 3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580
Qy 2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db 3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640
Qy 2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db 3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700
Qy 2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
Db 3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760
Qy 2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
Db 3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820
Qy 2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
Db 3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880
Qy 2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db 3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 8
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

Query Match 98.0%; Score 2414; DB 14; Length 3930;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```
Qy      14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy      74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy     134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy     194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy     254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy     314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy     374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy     434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy     494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy     554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy     614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy     674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy     734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy     794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy     854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy     914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy     974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy    1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy    1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy    1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy    1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy    1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy    1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
           |||||||||||||||  ||  |||||||||||||||||||||||||||||||||||||||
Db    2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy    1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy    1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy    1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy    1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy    1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy    1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220
```

```
Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 9
US-10-190-435-11
; Sequence 11, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna_C
US-10-190-435-11
```

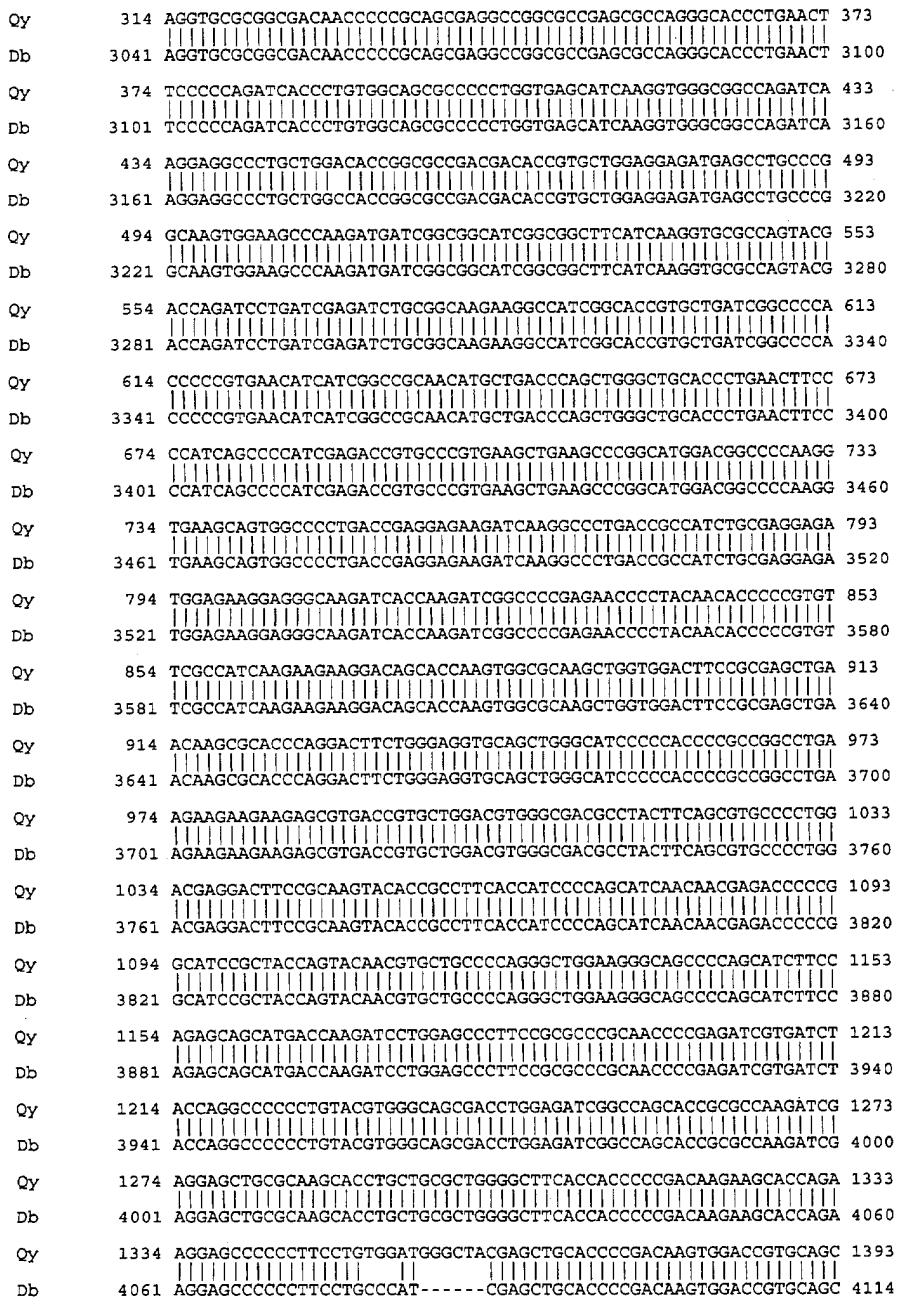
```
Query Match           98.0%;  Score 2414;  DB 14;  Length 3930;
Best Local Similarity  99.5%;  Pred. No. 0;
Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGGCCACATCGCCCGCA 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
```

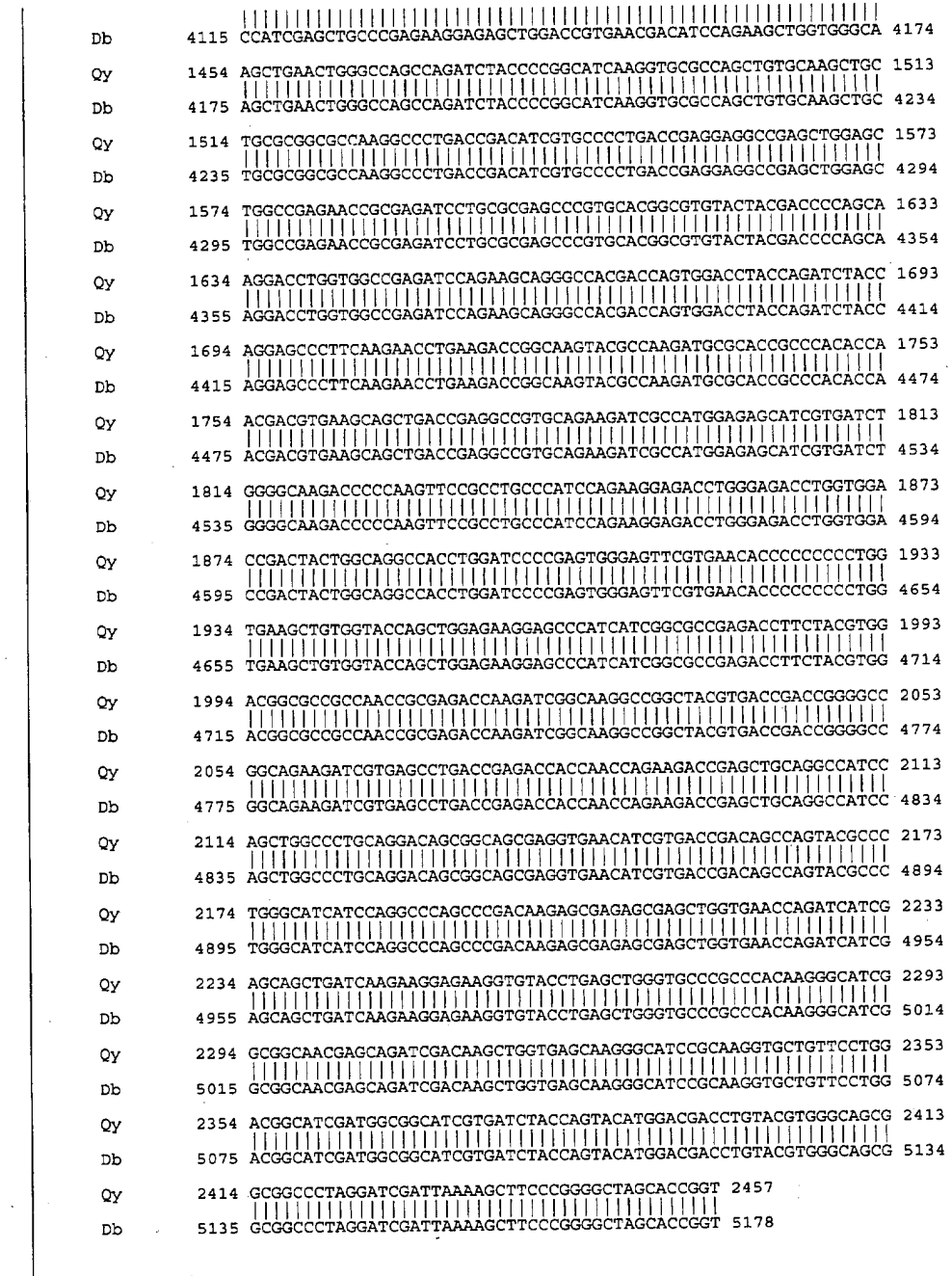
```
Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             |||||||||||||||    ||       ||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 10
US-10-190-435-58
; Sequence 58, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 5184
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C
US-10-190-435-58

Query Match 98.0%; Score 2414; DB 14; Length 5184;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040
```


Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy 1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db 3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000
Qy 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db 4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060
Qy 1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db 4061 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 4114
Qy 1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453


Db 4115 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 4174
Qy 1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db 4175 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 4234
Qy 1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db 4235 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 4294
Qy 1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db 4295 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 4354
Qy 1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db 4355 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 4414
Qy 1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db 4415 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 4474
Qy 1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db 4475 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 4534
Qy 1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
Db 4535 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 4594
Qy 1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db 4595 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 4654
Qy 1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db 4655 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 4714
Qy 1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db 4715 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 4774
Qy 2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db 4775 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 4834
Qy 2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db 4835 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 4894
Qy 2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db 4895 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 4954
Qy 2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
Db 4955 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 5014
Qy 2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
Db 5015 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 5074
Qy 2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
Db 5075 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 5134
Qy 2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db 5135 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178

RESULT 11

```
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: TatRevNefgagCpolIna C
US-10-190-305A-82

  Query Match             98.0%;  Score 2414;  DB 14;  Length 5184;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
              |||||||||||||||  ||           ||||||||||||||||||||||||||||||
Db       4061 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 4114

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4115 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 4174

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4175 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 4234

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4235 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 4294

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4295 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 4354

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4355 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 4414

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4415 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 4474

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4475 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 4534
```

```
Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4535 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 4594

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4595 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 4654

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4655 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 4714

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4715 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 4774

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4775 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 4834

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4835 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 4894

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4895 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 4954

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4955 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 5014

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5015 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 5074

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5075 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 5134

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||
Db       5135 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 12
US-10-190-435-13
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 3531
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13

Query Match 96.8%; Score 2383.6; DB 14; Length 3531;
Best Local Similarity 98.8%; Pred. No. 0;

Matches 2414; Conservative 0; Mismatches 24; Indels 6; Gaps 1;

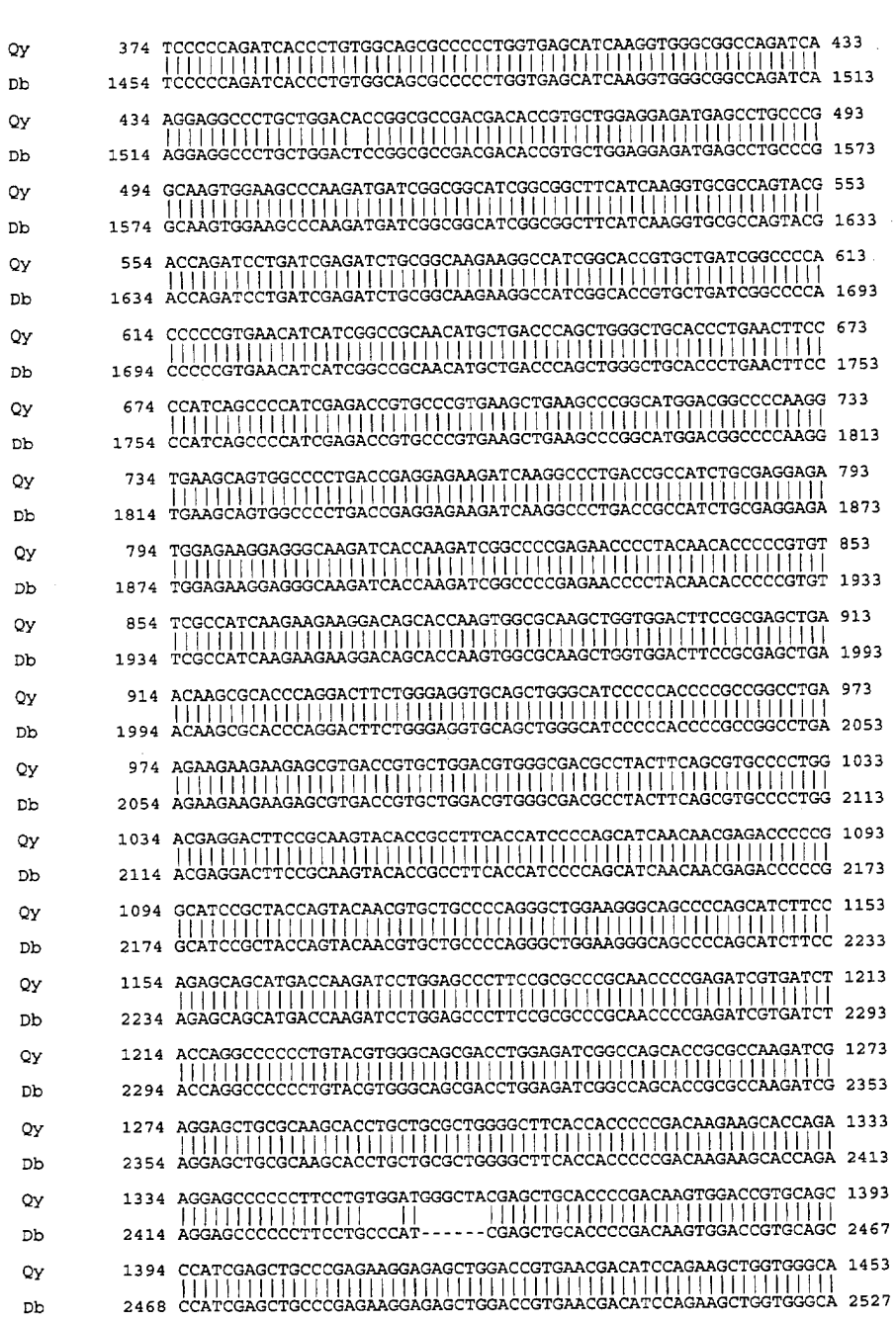
```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| |||||||||||| || |  |  ||  ||  | || ||| |     |
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2048 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2107

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2108 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2167
```

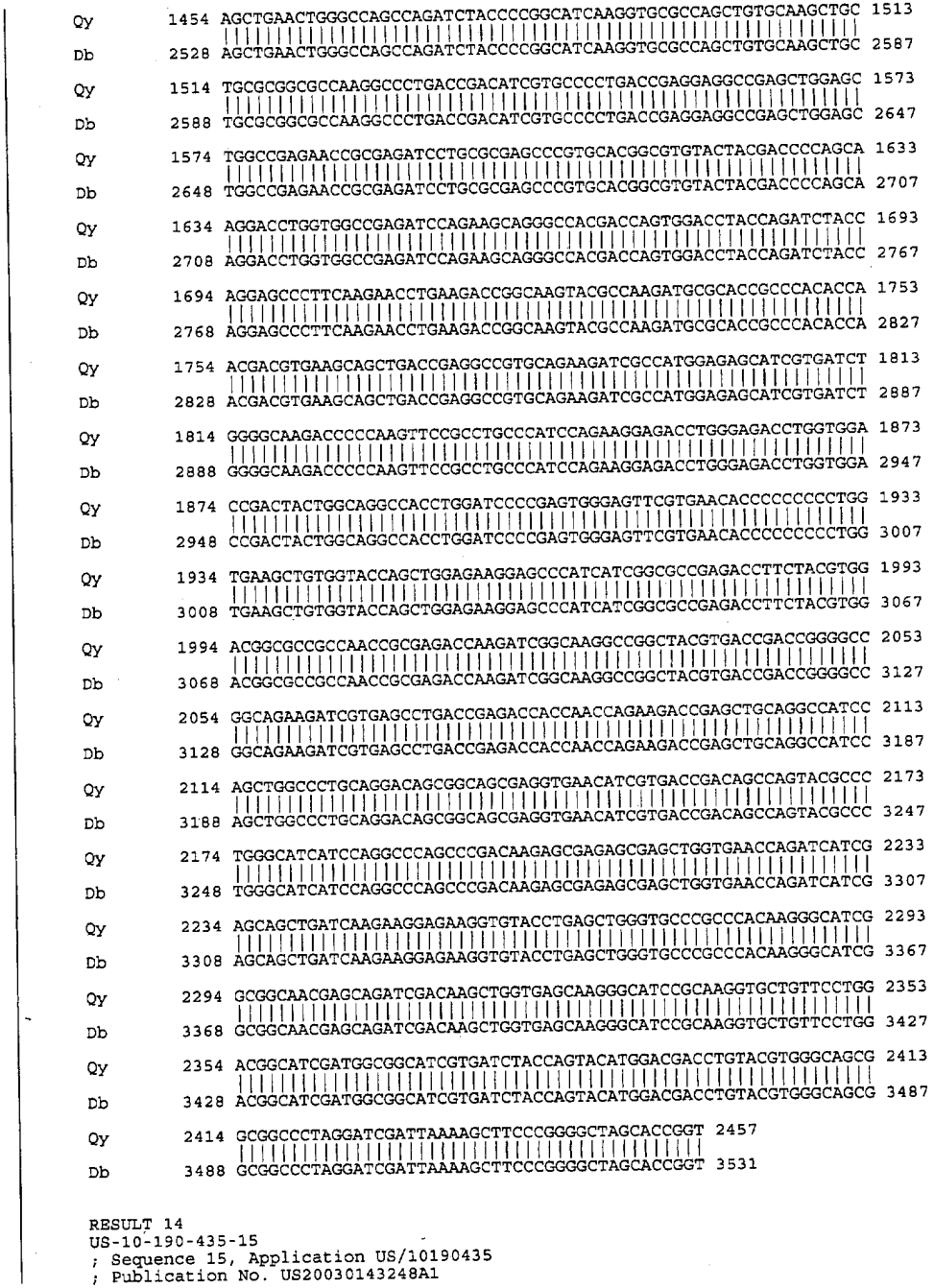
```
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2168 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCCAGCATCTTCC 2227

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCCGCAACCCCGAGATCGTGATCT 2287

Qy       1214 ACCAGGCCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2288 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2347

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2348 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2407

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             |||||||||||||||||  ||      ||||||||||||||||||||||||||||||||||||
Db       2408 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2461

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2462 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2521

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2522 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2581

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2582 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 2641

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2642 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 2701

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2702 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 2761

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2762 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 2821

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2822 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 2881

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2882 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 2941

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2942 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3001

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3002 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3061

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3062 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3121

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3122 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3181

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3182 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3241

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3242 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3301

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3302 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3361

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3362 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3421

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3422 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3481

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||
Db       3482 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3525
```

RESULT 13
US-10-190-435-14
; Sequence 14, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 3537
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmutAtt_C
US-10-190-435-14

Query Match 96.7%; Score 2382; DB 14; Length 3537;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2413; Conservative 0; Mismatches 25; Indels 6; Gaps 1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| |||||||||||| || |  |  ||  ||  || ||| |     |
Db       1094 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1153

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1154 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1213

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1214 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1273

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1274 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1333

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1334 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1393

Qy        314 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1394 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1453
```





RESULT 14
US-10-190-435-15
; Sequence 15, Application US/10190435
; Publication No. US20030143248A1

```
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 3537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagPolmutIna_C
US-10-190-435-15

  Query Match             96.7%;  Score 2382;  DB 14;  Length 3537;
  Best Local Similarity   98.7%;  Pred. No. 0;
  Matches 2413;  Conservative    0;  Mismatches   25;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1094 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1153

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1154 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1213

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1214 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1273

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       1274 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1333

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       1334 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1393

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db       1394 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1453

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db       1454 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1513

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db       1514 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1573

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db       1574 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1633

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db       1634 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1693

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db       1694 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1753

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db       1754 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1813

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db       1814 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1873

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db       1874 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1933

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db       1934 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1993

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db       1994 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2053

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       2054 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2113

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       2114 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2173

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db       2174 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2233

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db       2234 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2293

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db       2294 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2353

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db       2354 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2413

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db       2414 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2467

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
Db       2468 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2527

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db       2528 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2587

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db       2588 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 2647

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db       2648 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 2707

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db       2708 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 2767

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db       2768 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 2827

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db       2828 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 2887

Ov       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
```

```
Db       2888 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 2947
Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db       2948 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3007
Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db       3008 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3067
Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db       3068 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3127
Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db       3128 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3187
Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db       3188 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3247
Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db       3248 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3307
Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
Db       3308 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3367
Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
Db       3368 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3427
Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
Db       3428 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3487
Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db       3488 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3531
```

RESULT 15
US-10-190-435-12
; Sequence 12, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagComplPolmutInaTatRevNef_C
US-10-190-435-12

Query Match 96.7%; Score 2381; DB 14; Length 5145;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2400; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606
Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
```

```
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             |||||||||||||||||  ||      ||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGG 2424
             ||||||||||||
Db       3881 GCGGCCCTAGG 3891
```

Search completed: April 10, 2004, 21:24:41
Job time : 602.333 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:34:42 ; Search time 4148.08 Seconds
(without alignments)
17774.420 Million cell updates/sec

Title: US-09-610-313-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	90.8	3.7	2598	11	AY103647	AY103647 Zea mays
	2	85.6	3.5	951	12	BM321451	BM321451 rockefell
c	3	85	3.4	869	14	CK159167	CK159167 FGAS04056
	4	80.2	3.2	1132	12	BM320864	BM320864 rockefell
	5	79.6	3.2	1165	12	BM320900	BM320900 rockefell
	6	79	3.2	867	12	BM321430	BM321430 rockefell
	7	76.6	3.1	1550	12	BM321022	BM321022 rockefell
	8	75.6	3.1	671	13	CA093222	CA093222 SCCCCL300
	9	75.4	3.1	2299	11	AY106831	AY106831 Zea mays
	10	75	3.0	862	12	BM321023	BM321023 rockefell
	11	74.6	3.0	914	28	BZ568300	BZ568300 pacs2-164
	12	74.4	3.0	3134	11	AY109500	AY109500 Zea mays
c	13	73.8	3.0	889	14	CK159613	CK159613 FGAS04107
c	14	73.6	3.0	757	29	CC678788	CC678788 OGWII71TV
	15	73.2	3.0	853	12	BM321393	BM321393 rockefell
c	16	72.8	2.9	566	12	BM587428	BM587428 170006873
	17	72.6	2.9	788	14	CB643171	CB643171 OSJNEb03L
	18	72.4	2.9	753	29	CC675888	CC675888 OGWCO51TH
c	19	72.4	2.9	801	12	BJ375394	BJ375394 BJ375394
	20	72.2	2.9	688	14	CB648640	CB648640 OSJNEb12C
	21	72.2	2.9	764	14	CB651670	CB651670 OSJNEb16L
	22	72.2	2.9	766	14	CB642928	CB642928 OSJNEb03F
	23	72.2	2.9	809	14	CB641397	CB641397 OSJNEb01A
	24	72.2	2.9	841	14	CB651502	CB651502 OSJNEb16H
c	25	72.2	2.9	851	29	CG260623	CG260623 OGWGE95TV
c	26	72.2	2.9	860	29	CC691066	CC691066 OGUDF69TV
	27	71.8	2.9	1933	11	AY325173	AY325173 Rattus no
	28	71.8	2.9	5622	29	AY400359	AY400359 Homo sapi
	29	71.2	2.9	646	12	BM645449	BM645449 170006873
	30	71.2	2.9	724	12	BM584191	BM584191 170006872
	31	71	2.9	725	14	CB630209	CB630209 OSIIEb07A
	32	71	2.9	745	14	CB641936	CB641936 OSJNEb01N
	33	71	2.9	766	14	CB629976	CB629976 OSIIEb06K
	34	71	2.9	787	14	CB658685	CB658685 OSJNEc14O
	35	71	2.9	792	14	CB618305	CB618305 OSIIEa02B
	36	71	2.9	810	14	CB618374	CB618374 OSIIEa02D
	37	71	2.9	823	14	CB627740	CB627740 OSIIEb02O
	38	71	2.9	865	14	CB628660	CB628660 OSIIEb04F
c	39	71	2.9	925	29	CNS0091P	AL053013 Drosophil
	40	70.6	2.9	703	14	CB675843	CB675843 OSJNEe11M
	41	70.6	2.9	710	14	CB676048	CB676048 OSJNEe12B
	42	70.6	2.9	712	14	CB647546	CB647546 OSJNEb10I
	43	70.6	2.9	714	14	CB651194	CB651194 OSJNEb16B
	44	70.6	2.9	715	14	CB646590	CB646590 OSJNEb09C
	45	70.6	2.9	715	14	CB669855	CB669855 OSJNEe02K

ALIGNMENTS

RESULT 1
AY103647
LOCUS AY103647 2598 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2598)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2598)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES Location/Qualifiers
source 1. .2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

ORIGIN

Query Match 3.7%; Score 90.8; DB 11; Length 2598;
Best Local Similarity 43.3%; Pred. No. 0.21;
Matches 719; Conservative 0; Mismatches 909; Indels 33; Gaps 5;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db        351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410
Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db        411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470
Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCG 349
Db        471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530
Qy        350 CCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGA 409
Db        531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590
Qy        410 GCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCG 469
Db        591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650
Qy        470 TGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCG 529
Db        651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710
Qy        530 GCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCA 589
Db        711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767
Qy        590 TCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCC 649
Db        768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827
Qy        650 AGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGA 709
Db        828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887
Qy        710 AGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGG 769
Db        888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCTGTCCGCCGTCTTCTGCGAGG 947
Qy        770 CCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCG 829
Db        948 TCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACCCGG 1007
Qy        830 AGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCA 889
Db       1008 GGTCCATCGAGGCCGCGGCCATCATGGAGCACATCCTGGATGGCAGCTCCTTCATGAAGC 1067
Qy        890 AGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGG 949
```

```
Db       1068 AGGCCAAGAAGTTGAACGAGCTGGACCCGC----------TGCTGAAGCCCAAGCAGG 1115
Qy        950 GCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCG 1009
Db       1116 ACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTGGGCCCCCAGATCGAGGTCATCCGCG 1175
Qy       1010 ACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCC 1069
Db       1176 CCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTCATCGACG 1235
Qy       1070 CCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCT 1129
Db       1236 TCCACCGCGGCAAGGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTGTCCA 1295
Qy       1130 GGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCG 1189
Db       1296 TGGACAACGCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCG 1355
Qy       1190 CCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGA-- 1247
Db       1356 AGCTCGTCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGCAACC 1415
Qy       1248 ----CCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1303
Db       1416 CCAGCCTGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCCGAGC 1475
Qy       1304 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1363
Db       1476 TCCAGTACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGAGCAGCACAACCAGG 1535
Qy       1364 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1423
Db       1536 ACGTGAACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGGCGATCGACATCCTGA 1595
Qy       1424 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1483
Db       1596 AGCTCATGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCACCTCG 1655
Qy       1484 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1543
Db       1656 AGGAGAACATCAAGGCGTCGGTGAAGAACACCGTGACCCAGGTGGCCAAGAAGGTGCTGA 1715
Qy       1544 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAA---------CCGCG 1594
Db       1716 CCATGAACCCCTCGGGCGAGCTCTCCAGCGCCCGCTTCAGCGAGAAGGAGCTGATCAGCG 1775
Qy       1595 AGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCG 1654
Db       1776 CCATCGACCGCGAGGCCGTGTTCACGTACGCGGAGGACGCGGCCAGCGGCAGCCTGCCGC 1835
Qy       1655 AGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGA 1714
Db       1836 TGATGCAGAAGCTGCGCGCCGTGCTGGTGGACCACGCCCTCAGCAGCG---GCGACGCGG 1892
Qy       1715 ACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGC 1774
Db       1893 AGCGGGAGCCCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGC 1952
Qy       1775 TGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTG 1815
Db       1953 TGCCCCAGGAGGTGGAGGCCGCCCGCGTGGCGGTCGCCGAG 1993
```

RESULT 2
BM321451
LOCUS BM321451 951 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.
ACCESSION BM321451
VERSION BM321451.1 GI:18055857
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi

```
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 951)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 951   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .951
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.5%;  Score 85.6;  DB 12;  Length 951;
  Best Local Similarity   46.0%;  Pred. No. 0.73;
  Matches  323;  Conservative    0;  Mismatches  376;  Indels    3;  Gaps    1;

Qy       1734 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1793

Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1794 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1853

Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1854 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1913

Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1914 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1973

Db        399 CGACGGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1974 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2033

Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2034 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2093

Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2094 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2153

Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2154 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2213

Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy       2214 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2273

Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2274 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2333

Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2334 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393

Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2394 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2435

Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917


RESULT 3
CK159167/c
LOCUS       CK159167                 869 bp    mRNA    linear   EST 05-DEC-2003
DEFINITION  FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA,
            mRNA sequence.
ACCESSION   CK159167
VERSION     CK159167.1  GI:38985053
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLt537  row: N  column: 23.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
                     (equal amount of cDNA pooled together before subtraction,
                     tester) and subtracted against genotype Norstar cold
                     hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                     cDNA (Clontech)priming and non-directional cloning"
ORIGIN

  Query Match             3.4%;  Score 85;  DB 14;  Length 869;
  Best Local Similarity   45.4%;  Pred. No. 0.85;
  Matches  304;  Conservative    0;  Mismatches  365;  Indels    0;  Gaps    0;

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639

Db        827 ATGGCGGCGGCCCGCACAACGAGGATCACCAACCAGGACGAGAACAACCACAACCACCAC 768

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699

Db        767 CACAACGACGACAACAGCGACACAACGACCACCAACAACACCAACAACAACAACAACAAC 708

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
```

```
Db        707 ACCAACAACAACAACAACAACAACAACAACAACGACAACCACCACCACAACCACAACAAC 648

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819

Db        647 AACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 588

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879

Db        587 AACAACAACAACAACAACAACAACAACAACAACAACGACAACAACAACAACAACAACAAC 528

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939

Db        527 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 468

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999

Db        467 AACAACAACAACAACAACAACAACAACGACGACGACGACGACAACGACGACGACGACGACGAC 408

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059

Db        407 GACAACAACGACAACAACGACAACAACAACAACGACAACAACAACAACAACAACAACAAC 348

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119

Db        347 GACCACAACGACAACGACAACAACAACAACAACAACAACAACAACGACAACAACCACAAC 288

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179

Db        287 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 228

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTG 1239

Db        227 AACAACAACAACAACAACGACGACAACGACGTCAACGATAACAACGACAACAACGACAAC 168

Qy       1240 GGCAGCGAC 1248

Db        167 AACGGCAAC 159
```

RESULT 4
BM320864
LOCUS BM320864 1132 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.
ACCESSION BM320864
VERSION BM320864.1 GI:18055270
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1132 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1132
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN

Query Match 3.2%; Score 80.2; DB 12; Length 1132;
Best Local Similarity 45.3%; Pred. No. 3.3;
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

```
Qy        139 CGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGAC 198

Db         79 CGCCGCGAGGGCAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAG 138

Qy        199 TGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCC 258

Db        139 TACAACAGCCCCAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTC 192

Qy        259 CGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG 318

Db        193 TGCCAGATCGCCTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCAC 252

Qy        319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCC 378

Db        253 GAGCTCACCCGCTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACT 312

Qy        379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438

Db        313 GGCCTGCTGCTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGT 372

Qy        439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498

Db        373 GTCAAGAAGGTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGT 432

Qy        499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558

Db        433 TTCAAGGCCCTGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCC 492

Qy        559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618

Db        493 GCCCTCAAGGGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTC 549

Qy        619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678

Db        550 GTCGGCTTCAACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTC 609

Qy        679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738

Db        610 GGCGGCCACGTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGAC 669

Qy        739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798

Db        670 CGCCAGTTCTCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTAC 729

Qy        799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858

Db        730 ACCGAGGCCCACAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAG 789

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918

Db        790 CCCGAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966

Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897
```

RESULT 5
BM320900
LOCUS BM320900 1165 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.
ACCESSION BM320900
VERSION BM320900.1 GI:18055306

```
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1165   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1165
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79.6;  DB 12;  Length 1165;
  Best Local Similarity   45.1%;  Pred. No. 4;
  Matches  428;  Conservative    0;  Mismatches  506;  Indels   16;  Gaps    3;

Qy         90 CATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCG 149
Db         46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTCCAGACCCAGTTCCGTCGCCGCCGCGAGGG 105

Qy        150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209
Db        106 CAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAGTACAACAGCCC 165

Qy        210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269
Db        166 CAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTCTGCCAGATCGC 219

Qy        270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329
Db        220 CTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCACGAGCTCACCCG 279

Qy        330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389
Db        280 CTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACTGGCCTGCTGCT 339

Qy        390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGA 449
Db        340 GGCCCGCCGTGTGCTGAAGAAGCTCAACCTCTACTCCAAGTACGAGGGTGTCAAGAAGGT 399

Qy        450 CACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAA 509
Db        400 CAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGTTCAAGGCCC-- 457

Qy        510 GATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGA 569
Db        458 --TGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCCGCCCTCAA 515

Qy        570 GATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCAT 629
Db        516 GGGCATGTGCGACGGCGGCGTCAACGTCCCCCACAGCGAGACCCGCTTCGTCGGCTTCAA 575

Qy        630 CGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGA 689
Db        576 CGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCACGT 635

Qy        690 GACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCT 749
Db        636 CGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGACCGCCAGTTCTC 695

Qy        750 GACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAA 809
Db        696 GCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTACACCGAGGCCCA 755

Qy        810 GATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAA 869
Db        756 ------CAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAGCCCGA 809

Qy        870 GGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGA 929
Db        810 GGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAGAACCG 869

Qy        930 CTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGT 989
Db        870 CGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCCAAGAGCCAGTAAATTCC 929

Qy        990 GACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGG 1039
Db        930 CGGTGCCCTGTACACCGCCATCTCCCGCTGCGGCGTTGCGCTGCTNGCGG 979


RESULT 6
BM321430
LOCUS       BM321430                 867 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
            sequence.
ACCESSION   BM321430
VERSION     BM321430.1  GI:18055836
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 867   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79;  DB 12;  Length 867;
  Best Local Similarity   45.7%;  Pred. No. 4.4;
  Matches  314;  Conservative    0;  Mismatches  370;  Indels    3;  Gaps    1;

Qy         58 CAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAG 117
Db        131 CCGCACAAGATGCGCGAGTGCCTGCCGGTCATCATCCTGCTGCGCAACAGGTTGAAGTAC 190

Qy        118 GGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAG 177
```

```
Db        191 GCGCTGACCCGCCGTGAGGTCACCTCGATCGTGATGCAGCGCCTGATCAAGATCGACGGC 250

Qy        178 GAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237

Db        251 AAGGTCCGCACCGACACGCACTTCCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310

Qy        238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCC 297

Db        311 ACCGACGAGCACTTCCGCCTGCTCTACGACACCAAGGGCCGCTTCCAGGCGCACCGCATC 370

Qy        298 ACCAGCCGCGAGCTGCAGGT---GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354

Db        371 AACTCGACCGAGGCCAAGTTCAAGCTCGGCAAGGTCCGCCGCGTGCAGCTCGGCAACAAG 430

Qy        355 CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATC 414

Db        431 GGCATCCCGTACCTGGTGACCCACGACGCGCGCACGATCCGCTACCCCAACCCCGACATC 490

Qy        415 AAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTG 474

Db        491 AAGGTCAACGACACGGTCAAGATCGACCTGCGCTCGGGCAAGATCATCGACTTCGTCAGG 550

Qy        475 GAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTC 534

Db        551 TTCGAGATCGGCAACCTCGTCATGATCACTGGCGGACGCAACCTTGGCCGCGTCGGCGTC 610

Qy        535 ATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGC 594

Db        611 ATTGTGCGCCGCGAGAAGCACGAGGGCTCGTTCGAGATCATCCACGTCAAGGACGCCGTC 670

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654

Db        671 GGCCACCAGTTCGCGACGCGCCTGACCAACGTCTTCGTGATCGGCAAGGGCACCAAGTCC 730

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714

Db        731 CTCGTCACGCTGCCCGCCGGCAAGGGCATCAAGAAGTCGATCATCGAGGAGTTCCAGGCG 790

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAG 741

Db        791 CGCCACGCGCACAGGACCGAGCAGGAG 817
```

RESULT 7
BM321022

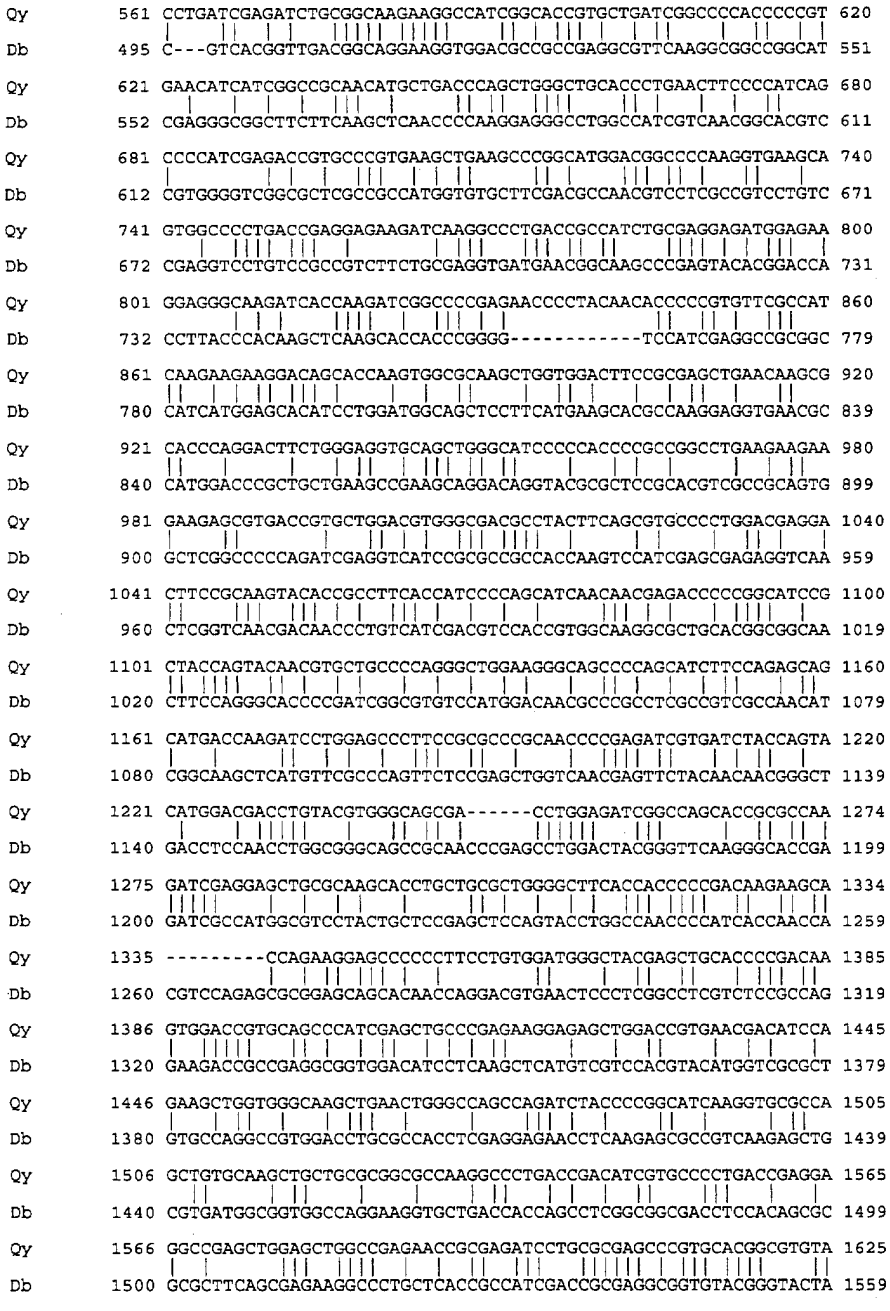
```
LOCUS       BM321022                1550 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1192 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321022
VERSION     BM321022.1  GI:18055428
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1550)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1550   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1550
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN
```

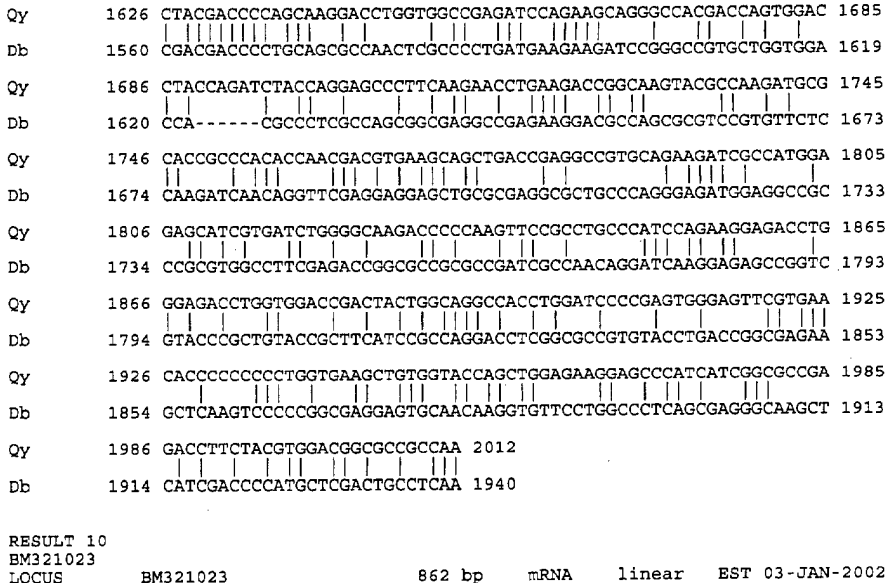
```
  Query Match             3.1%;  Score 76.6;  DB 12;  Length 1550;
  Best Local Similarity   45.6%;  Pred. No. 9.5;
  Matches  308;  Conservative    0;  Mismatches  364;  Indels    3;  Gaps    1;

Qy       1761 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1820

Db          5 CGGCACGAGGACGCAGGACCACGCGGCCGCCGCCATCGCGCAGGCCGGCGTCTCGGTCTT 64

Qy       1821 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1880

Db         65 CGCCTGGAAGGGCGAGAACCTCCAGGAGTACTGGGAGTGCACCTGGAAGGCCCTGTGCTT 124

Qy       1881 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1940

Db        125 CGGCCCCTACCAGGGCCCTCAGATCATCGTCGACGACGGCGGTGACGCGACTCTGATGAT 184

Qy       1941 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 2000

Db        185 CCACAAGGGGTTCGCGGCCGAGGACAACCCCAAGCTGCTGGAGGACGACGAGGGCCTCGA 244

Qy       2001 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2060

Db        245 GGAGGTCGCCTGCCTCAACAACGTGCTCAAGCAGGTCCAGAAGGAGCAGCCCGGCTTCTG 304

Qy       2061 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2120

Db        305 GCACAAGATCCTCCCCGAGATCCGCGGTGTCAGCGAGGAGACGACGACTGGCGTGATGAG 364

Qy       2121 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2180

Db        365 GCTGTACCAGCTGCACCGCGACGGCAAGCTGCTGTTCCCGGCCGTCAACGTCAACGACTC 424

Qy       2181 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2240

Db        425 TG---TCACCAAGAGCAAGTTTGACAACATCTACGGCTGCCGCCACTCGCTCATCGACGG 481

Qy       2241 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2300

Db        482 CATCAAGCGCGGCACCGACGTGATGCTCGGCGGCAAGGTCGCCGTCGTCGCGGGCTACGG 541

Qy       2301 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2360

Db        542 CGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCGCGGCCAGGGCTGCCGCGTCATCGTGAC 601

Qy       2361 CGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGG 2420

Db        602 GGAGATCGACCCCATCTGCGCGCTGCAGGCGTCGATGGCGGGCTTCGAGGTCAACACGCT 661

Qy       2421 CGGCCCTAGGATCGA 2435

Db        662 CGAGGCGGGGCTCGA 676
```

RESULT 8
CA093222

```
LOCUS       CA093222                671 bp    mRNA    linear   EST 23-SEP-2003
DEFINITION  SCCCCL3001F12.g CL3 Saccharum officinarum cDNA clone SCCCCL3001F12
            5', mRNA sequence.
ACCESSION   CA093222
VERSION     CA093222.1  GI:34946529
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 001  row: F  column: 12
            Seq primer: M13/Reverse primer.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCCCCL3001F12"
                     /lab_host="XL1Blue MRF'"
                     /clone_lib="CL3"
                     /note="Organ: Pool of sugarcane calli submitted to low
                     (4oC) and high (37 C) temperature stress; Vector:
                     pBlurScript; Site_1: EcoRI; Site_2: XhoI; An
                     unidirectional cDNA library generated from [Pool of
                     sugarcane calli  submitted to low (4oC) and high (37 C)
                     temperature stress]. cDNA was  prepared from polyA+ mRNA
                     using ZAP - cDNA Synthesis Kit (Stratagene).  The
                     double-strand cDNAs were fractionated in a sepharose CL-2B
                     40cm-columns and fragments sizing between 0.8 and 1.5 Kb
                     were  directionally cloned into the vector. Details of
                     each source of RNA and  library construction can be
                     obtained at  http://sucest.lad.ic.unicamp.br/public"
ORIGIN

  Query Match             3.1%;  Score 75.6;  DB 13;  Length 671;
  Best Local Similarity   47.3%;  Pred. No. 11;
  Matches  228;  Conservative    0;  Mismatches  254;  Indels    0;  Gaps    0;

Qy        986 GCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCC 1045
              || | | || | |  |||  | || |||    |||| |||| ||||   | ||  |||| |
Db        125 GCCTCAGCGAGGTCTACGGCGCCGCCGCGGCATTCGGCGGGCCCAATGCCGGTGACTACG 184

Qy       1046 GCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACC 1105
              ||     | |  |  |||    |  |  |    ||| | | |  | |   ||    ||
Db        185 GCGGCGAGATGGGGTTCCTTGACGTCGTCGAGCCCAAGGCGGCGGCGGCGCTCGATGACG 244

Qy       1106 AGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGA 1165
                | || |   | |  || ||| ||||     ||| || |  |  ||||| ||
Db        245 GCGTCGCCGCGGGCCTCGGGGTCTGCAAGGTGGAGCCCGGCCTAGTGGAGAGCGGCGGTG 304

Qy       1166 CCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGG 1225
              |      ||  |  ||| ||  |||  |      | |      | |  | |  || | | ||
Db        305 CGTTTGGCCCTGGACCCGTCGTCGCGCCGCCGGCGGCCCCGGCGTCCAAGAAGAAGAGGG 364

Qy       1226 ACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGC 1285
               |||      |  ||   ||||  |  ||  |  | |||| ||||   |    |
Db        365 TCGAGGGGATGCCGTCCAAGAACCTGATGGCCGAGCGCCGCCGCCGCAAGCGCCTCAACG 424

Qy       1286 TGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGC 1345
               |||    | |||||||||  | | |  |||    |  ||||| | ||| || |  |
Db        425 ACCGCCTCTCCATGCTGCGCTCCGTCGTGCCCAAGATCAGCAAGATGGACCGGACGTCGA 484

Qy       1346 CCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1405
               ||     |    ||| | ||||  |  | | |     || ||| |  || ||     |
Db        485 TCCTGGGCGACACGATCGACTACATGAAGGAGCTGCTGGAGAGGATCAAGCTGCTGCATG 544

Qy       1406 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1465
              || |  |||  || ||     |  | |  | ||| |  ||  |  |  |  |||| |
Db        545 AGGAGATCGACCAGCAGCAGCAGGAGGGGCCGGGCATGCTCAACGTCTTCCGGGAGCTCA 604

Qy       1466 AC 1467
              ||
Db        605 AC 606


RESULT 9
AY106831
LOCUS       AY106831               2299 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO142079 mRNA sequence.
ACCESSION   AY106831
VERSION     AY106831.1  GI:21209909
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2299)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2299)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .2299
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638374"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match             3.1%;  Score 75.4;  DB 11;  Length 2299;
  Best Local Similarity   43.0%;  Pred. No. 14;
  Matches  734;  Conservative    0;  Mismatches  936;  Indels   37;  Gaps    6;

Qy        321 CGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCA 380
              |||| | ||  ||  ||     |  ||  |    |  |  || ||  | |||| |
Db        256 CGGCTCCGACGGCCACACGCTCCCGTCCGAGGTTTCGCGTGCGGCCATGCTCGTCCGCAT 315

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
              | |||||||    ||| ||  |  |  | |  |  | |  | |||  ||
Db        316 CAACACCCTCCTCCAGGGCTACTCCGGCATCCGCTTCGAGATCCTCGAGGCCATCACCAA 375

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
              ||  | |||||||||| |  |    || |||  |  |  |  | | |  |||     ||
Db        376 GCTCATCAACACCGGCGTCAGCCGTGCC-TGCCGCTGCGGGGCACCATCACCGCGTCGGG 434

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
              | | |    |  |  |  ||||| |||| |||||| || |||  |   |||  |||
Db        435 CGACCTCGTCCCGCTGTCCTACATCGCCGGCCTCATCACGGGGCGTCCCAACGCGCAGGC 494
```





```
RESULT 10
BM321023
LOCUS       BM321023                862 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.594 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321023
VERSION     BM321023.1  GI:18055429
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 862)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 862   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .862
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.0%;  Score 75;  DB 12;  Length 862;
  Best Local Similarity   46.2%;  Pred. No. 13;
  Matches  283;  Conservative    0;  Mismatches  327;  Indels    3;  Gaps    1;

Qy       1734 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1793
```

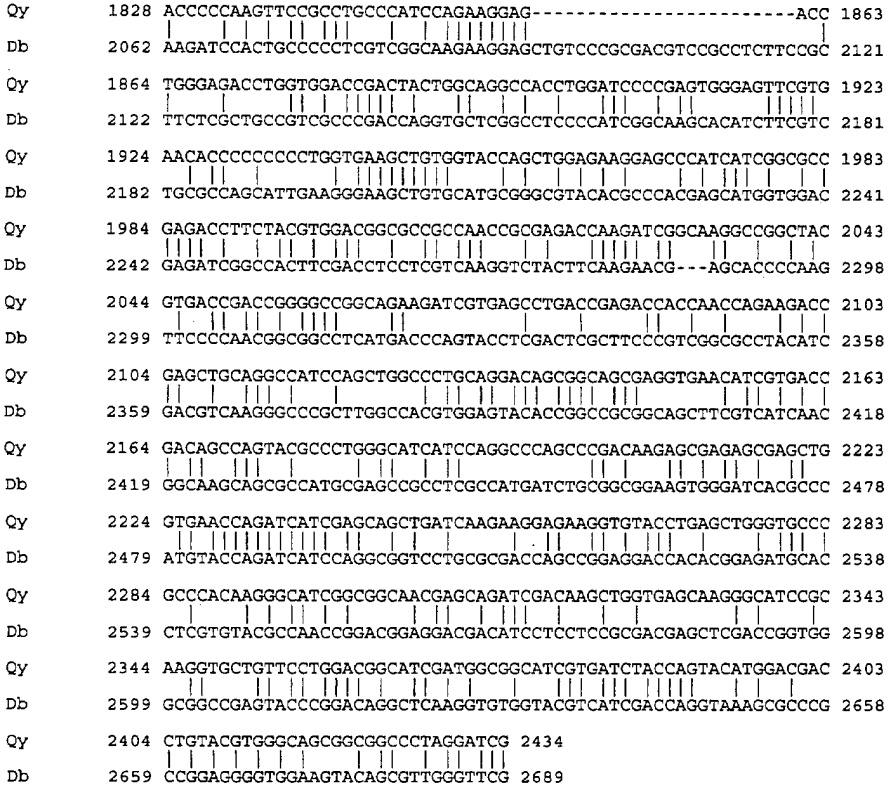
```
Db        220 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 279

Qy       1794 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1853

Db        280 CATCGCGCAGGCCGGAGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 339

Qy       1854 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1913

Db        340 GGAGTGCACCTNGAAGGCCCTGTGCTTCGGCCCCTACCAGGGACCTCAGANCATCGTCGA 399

Qy       1914 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1973

Db        400 CGACGGCGGTGACGCGACTCTAATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 459

Qy       1974 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2033

Db        460 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 519

Qy       2034 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2093

Db        520 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 579

Qy       2094 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2153

Db        580 CGAGGAGACGACGACTGGCGTGATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 639

Qy       2154 CATCGTGACCGACAGCCAGTACGCCCTGGGCATC---ATCCAGGCCCAGCCCGACAAGAG 2210

Db        640 GTTCCCGGCCGTCAACGTCAACGACTCTGTCACCAAGAGCAAGTTTGACAACATCTACGG 699

Qy       2211 CGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCT 2270

Db        700 CTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGGCAA 759

Qy       2271 GAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAG 2330

Db        760 GGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCGCGG 819

Qy       2331 CAAGGGCATCCGC 2343

Db        820 CCAGGGCTGCCGC 832
```

```
RESULT 11
BZ568300
LOCUS       BZ568300                914 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_7463.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_7463, genomic survey sequence.
ACCESSION   BZ568300
VERSION     BZ568300.1  GI:27201058
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .914
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_7463"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN
```

```
  Query Match             3.0%;  Score 74.6;  DB 28;  Length 914;
  Best Local Similarity   46.8%;  Pred. No. 15;
  Matches  270;  Conservative    0;  Mismatches  304;  Indels    3;  Gaps    1;

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711

Db        112 CTGCGCAACCAGCAGGACGGCCTGATCTTCAGCCAGGCCGGCGCCCTCGAGGTGCAGGCC 171

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771

Db        172 GGCAGCCTGGACAACCGCCAGGGCACGCTCCAGGCCCAGGGCGACAACCAGCTGCGTATC 231

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831

Db        232 GGTGGCGCGCTGGACAACCAGGGCGGCCGCCTGGACAGCCGGGCCGGTAACCTAGACCTG 291

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891

Db        292 CAGAGCGGCAGCCTCGACAACGGCGCCGGCGGCGTGCTCAACAGCGCCAAGGGTTGGCTG 351

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951

Db        352 AAGCTGGTCACCGGGCTGTTCGACAACAGCGCCGGCGTCACCCAGGCGCAGTCGCTGGAA 411

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011

Db        412 ATCCGCGCCGGGCAAGGCGTGCGCAACCAGCAGGGC---CATGTCTCGGCGCTGGGCGGC 468

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071

Db        469 GACAACCGCATCGTCACCGTCGACTTCGACAACCAGGGTGGCGGCCTCTACGCCAGCGGC 528

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131

Db        529 CTGCTCAGCCTCGACGGCCAGCGCTTCCTCAACCAGGGCGCGGCGGCGGGCCAGGGCGGC 588

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191

Db        589 AAGGTCGGCGCCGGGCGCATCGACTTCAGCCTGGCCGGCGCGCTGGCCAACCGCTTCGCC 648

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACG 1228

Db        649 AGTTGGAAAGCGAGAGCGAGCTGCACCTGCGCGCCCG 685
```

```
RESULT 12
AY109500
LOCUS       AY109500                3134 bp    mRNA    linear   HTC 17-OCT-2002
DEFINITION  Zea mays CL1506_1 mRNA sequence.
ACCESSION   AY109500
VERSION     AY109500.1  GI:21213244
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3134)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 3134)
  AUTHORS   Coe,E.H.
```

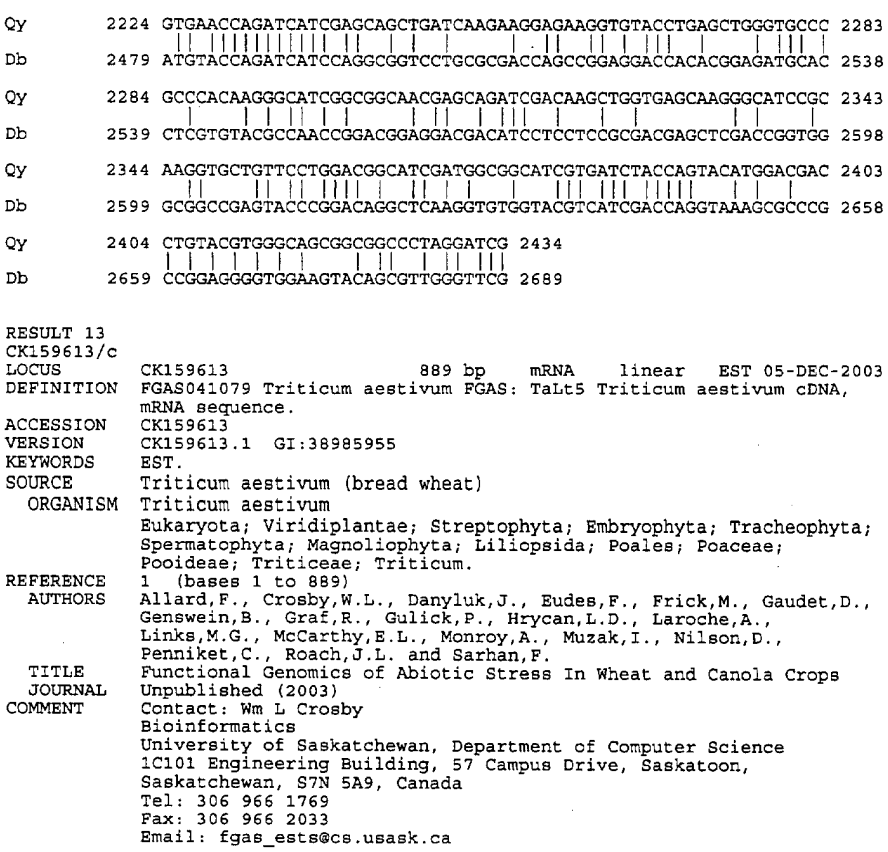
```
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .3134
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:630364"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match             3.0%;  Score 74.4;  DB 11;  Length 3134;
  Best Local Similarity   40.6%;  Pred. No. 20;
  Matches  963;  Conservative    0;  Mismatches 1347;  Indels   61;  Gaps    8;

Qy        115 GAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGC 174
Db        329 GAGAGCTGTACGGCTCGCACCTGCAGCTGGAGGTGGAGCCCGCGGTGCAGGACGCGCGCG 388

Qy        175 AAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGAC 234
Db        389 ACGAAGGCACCGCGGACGCGTGGATCGAGCGGAACCCGTGCCTGGTGAGGCTGACGGGGA 448

Qy        235 CTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGC 294
Db        449 AGCACCCGCTCAACTGCGAGCCGCCGCTAGCGCGGCTG-ATGCACCACGGCTTCATCACC 507

Qy        295 CCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354
Db        508 CCGGCGCCGCTGCACTACGTGCGCAACCACGGCGCCGTGCCGCNNNNNNACTGGGCGACG 567

Qy        355 CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATC 414
Db        568 TGGACCGTGGAGGTGACGGGCCTCGTGAGGCGGCCCGCGCGGCTCACCATGGAGGAGCTG 627

Qy        415 AAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTG 474
Db        628 GCGCGCGACTTCCCCGCCGTGGAGATCCCCGTGACGCTGGCGTGCGCGGGCAACCGGCGC 687

Qy        475 GAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTC 534
Db        688 AAGGAGCAGAACATGGTGCAGCAGACGGTGGGCTTCAACTGGGGCGCCGCCGGCGTGTCC 747

Qy        535 ATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGC 594
Db        748 ACGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGTGCTCCGCCGC---TGCGGCATCGTG 804

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654
Db        805 CCCCGCAAGGGCGGCGCCCTCAACGTGTGCTTCGAGGGCGCCGAGGNNNNNNNNNNNNN 864

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714
Db        865 NNNNNNNNNNGTACGGCACCAGCGTCACGCGCGAGTGGGCCCTGGACCCGTCGCGGGAC 924

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTG 774
Db        925 ATCATGCTCGCCTACATGCAGAACGGCGAGCCGCTGCTGCCGGACCACGGCTTCCCCGTG 984

Qy        775 ACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAAC 834
Db        985 CGCGTCATCATCCCCGGCTGCATCGGTGGCCGCATGGTCAAGTGGCTCAAGCGC---ATC 1041

Qy        835 CCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTG 894
Db       1042 ATCGTCANNNNNGCCGAGTCCGACAATTACTACCATTTCAAGGACAACCGCGTCCTGCCG 1101

Qy        895 GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954
Db       1102 TCGCACGTCGACGCCGAGCTCGCCAACGCAGAAGCGTGGTGGTACAAGCCGGAGTACATC 1161

Qy        955 CCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC 1014
Db       1162 ATCAACGAGCTGAACATAAACTCGGTGATAACGACGCCGGGGCACGACGAGATCCTGCCC 1221

Qy       1015 TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGC 1074
Db       1222 ATCAACAGCATTACCACACAGCGCGGCTACACCATGAAAGGATACGCCTACTCCGGCGGC 1281

Qy       1075 ATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAG 1134
Db       1282 GGCAAGAAGGTGACGCGGGTGGAGGTGACGCTGGACGGCGGCGAGACATGGCTGGTGTGC 1341

Qy       1135 GGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGC 1194
Db       1342 CACCTCGACCACCCGGAGAAGCCCAACAAGTACGGCAAGTACTGGTGCTGGTGCTTCTGG 1401

Qy       1195 AACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAG 1254
Db       1402 TCCGTCGAGGTGGAGGTCCTCGACCTGCTCGGCGCCAAGGAGATCGCCGTGCCCGCATGG 1461

Qy       1255 ATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTC 1314
Db       1462 GACCAGTCGCTCAACACCCAGCCCGAGAAGCTCATATGGAACCTCATGGGGATGATGAAC 1521

Qy       1315 ACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTG 1374
Db       1522 AACTGCTGGTTCAAGGTGAAGGTGAACGTGTGCCGTCCGCACAAGGGCGAGATCGGGCTG 1581

Qy       1375 CACCCCGACAAGTGGACCGTGCAGCCCATCGAGC-----------TGCCCGAGAAGGAG 1422
Db       1582 GTGTTCGAGCACCCGACGCAGCCCGGCAACCAGCCCGGCGGGTGGATGGCGCGGCAGAAG 1641

Qy       1423 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1482
Db       1642 CACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAGCGCAGCACGTCCACGCCGTTC 1701

Qy       1483 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1542
Db       1702 ATGAACACCACCGACGTCGGCAAGCAGTTCACCATGTCCGAGGTGCGCAAGCACGCGTCG 1761

Qy       1543 GACATCGTGCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGG------CCGAGAACCGCGAG 1596
Db       1762 CAGGAGTCGGCGTGGATCGCCGTGCACGGCCACGTCTACGACTGCACCAAGTTCCTCAAG 1821

Qy       1597 ATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAG 1656
Db       1822 GACCACCCCGGCGGCGCCGACAGCATCCTCATCAACGCCGGCACCGACTGCACCGAGGAG 1881

Qy       1657 ATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC--------- 1707
Db       1882 TTCGACGCCATCCACTCCGACAAGGCCAAGGCGCTCCTCGACACCTACCGCATCGGCGAG 1941

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
Db       1942 CTCATCACCACGGGCACCGGCTACAGCTCCGACAACTCCGTCCACGGCGGCTCCGTCCTG 2001

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db       2002 TCGCACCTCGCGCCCATCCGCGAGGCCGTCAGGGNNNNNGCGCTCTCCAACCCGCGCGAT 2061
```



RESULT 13
CK159613/c
LOCUS CK159613 889 bp mRNA linear EST 05-DEC-2003
DEFINITION FGAS041079 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK159613
VERSION CK159613.1 GI:38985955
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 889)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_ests@cs.usask.ca

This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [123,668].
Plate: TaLt539 row: N column: 07.
FEATURES Location/Qualifiers
source 1. .889
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaLt5"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"
ORIGIN

Query Match 3.0%; Score 73.8; DB 14; Length 889;
Best Local Similarity 47.4%; Pred. No. 18;
Matches 251; Conservative 0; Mismatches 277; Indels 1; Gaps 1;



RESULT 14
CC678788/c
LOCUS CC678788 757 bp DNA linear GSS 19-JUN-2003
DEFINITION OGWII71TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0595K21, genomic survey sequence.
ACCESSION CC678788
VERSION CC678788.1 GI:32083564
KEYWORDS GSS.
SOURCE Zea mays

```
ORGANISM   Zea mays
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
           clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1  (bases 1 to 757)
  AUTHORS  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
           Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
           Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE    Consortium for Maize Genomics
  JOURNAL  Unpublished (2002)
COMMENT    Other_GSSs: OGWII71TH
           Contact: Cathy Whitelaw
           TIGR
           9712 Medical Center Drive, Rockville, MD 20850, USA
           Tel: 301-838-5843
           Fax: 301-838-0208
           Email: whitelaw@tigr.org
           Seq primer: TF
           Class: sheared ends.
FEATURES            Location/Qualifiers
     source         1..757
                    /organism="Zea mays"
                    /mol_type="genomic DNA"
                    /strain="B73"
                    /db_xref="taxon:4577"
                    /clone="ZMMBMa0595K21"
                    /clone_lib="ZM_0.7_1.5_KB"
                    /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                    methylation filtered genomic DNA library"
ORIGIN

  Query Match             3.0%;  Score 73.6;  DB 29;  Length 757;
  Best Local Similarity   46.6%;  Pred. No. 19;
  Matches  306;  Conservative    0;  Mismatches  344;  Indels    6;  Gaps    2;

Qy        978 GAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGA 1037
Db        754 GCAGGTGGGCGTGAGCGACGTGTTCGCCGTGCCGGGGGACTTCAACCTGACGCTGCTGGA 695

Qy       1038 GGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG---ACCCCCGG 1094
Db        694 CCACCTGATCGCGGAGCCCGGCCTCCGCGTGTTCGGCTGCTGCAACGAGCTCAACGCCGG 635

Qy       1095 CATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCA 1154
Db        634 GTACGCCGCCGACGGGTACGCGCGCGCGCGGGGCGTGGGCGCCTGCGTTGTCACCTTCAC 575

Qy       1155 GAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTA 1214
Db        574 CGTGGGCGGGCTCAGCGTGCTCAACGCCATCGCCGGCGCCTACAGCGAGAACCTGCCCGT 515

Qy       1215 CCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAA 1274
Db        514 GGTGTGCGTGGTGGGCGGGCCCAACTCCAACGACTACGGCACCAACCGCGTCCTCCACCA 455

Qy       1275 GATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCA 1334
Db        454 CACCATCGGCCTCCCGGACTTCTCGCAGGAGCTGCGGTGCTTCCAGCCCGTCACCTGCCA 395

Qy       1335 CCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGT 1394
Db        394 CCAGGCCGTCGTCAACAACCTGGAGGACGCGCACGA---GCAGATCGACGGGGCCATCTC 338

Qy       1395 GCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGT 1454
Db        337 CACGGCGCTCAGGGAGAGCAAGCCCGTCTACATGAGCGTCAGCTGCAACCTGCCGGGGCT 278

Qy       1455 GGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAA 1514
Db        277 GCCGCATCCGACGTTCGCCACGGACCCTGTGCCATACTTGCTGGCGCCAAGGCCTAGCAA 218

Qy       1515 GCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCT 1574
Db        217 CCTGACGGGTCTCGCCGCCGCCGTCGAGGCCACCGTGTCGTTCCTCGACAAGGCCGTGAA 158

Qy       1575 GGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACG 1630
Db        157 GCCGGTGGTGGTGGCCGGCCCCAAGATCCGCGTGGCCAAGGCCGCCGCCGCCTTCG 102


RESULT 15
BM321393
LOCUS       BM321393                 853 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1222 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM321393
VERSION     BM321393.1  GI:18055799
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 853   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..853
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.0%;  Score 73.2;  DB 12;  Length 853;
  Best Local Similarity   46.9%;  Pred. No. 22;
  Matches  299;  Conservative    0;  Mismatches  329;  Indels    9;  Gaps    2;

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db        125 CTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGTGTCAAGAAG 184

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db        185 GTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGCCCCGCCCGTTCAAGGCC 244

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        245 CTGCTCGACGTCGGCCTGGTCCGCACCTCGACCGGCGCCCGCGTGTTCGCCGCCCTCAAG 304

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        305 GGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTCGTCGGCTTC 361

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        362 AACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCAC 421

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        422 GTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGACCGCCAGTTC 481
```

```
Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |  | | |  |||       | |    ||| |||   |||| ||||  | |  |||| |
Db        482 TCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTACACCGAGGCC 541

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              |      ||||      || ||   ||||    |  |  ||||  |  ||||   |||
Db        542 CA------CAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAGCCC 595

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||| |  ||        || |  |      |  |   |      ||  ||||||   |
Db        596 GAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAGAAC 655

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              | ||    ||  | || |  || |   | ||||| || ||  ||||   || | |
Db        656 CGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCCAAGAGCCAGTAAATT 715

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCG 1024
              |  ||| ||   | |  | ||  || | |||
Db        716 CCCGGTGCCCTGTACACNGCCATCTCCCGCTGCGGCG 752
```

Search completed: April 10, 2004, 16:16:19
Job time : 4157.08 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:53:16 ; Search time 6342.08 Seconds
(without alignments)
16873.640 Million cell updates/sec

Title: US-09-610-313-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2469	100.0	2469	6	AX455914	AX455914 Sequence
2	2442.2	98.9	2463	6	AX455915	AX455915 Sequence
3	2415.4	97.8	2457	6	AX455916	AX455916 Sequence
4	2052	83.1	2312	6	BD263706	BD263706 Improved
5	2052	83.1	2312	6	AR373389	AR373389 Sequence
6	2051.2	83.1	9166	6	AX427930	AX427930 Sequence
7	2042.8	82.7	9788	6	AX427936	AX427936 Sequence
8	2028.8	82.2	9169	6	AX427931	AX427931 Sequence
9	2027	82.1	9194	6	AX427925	AX427925 Sequence
10	2027	82.1	12411	6	AX427927	AX427927 Sequence
11	2025.4	82.0	9194	6	AX427926	AX427926 Sequence
12	2025.2	82.0	2306	6	BD263704	BD263704 Improved
13	2025.2	82.0	2306	6	AR373387	AR373387 Sequence
14	2020.4	81.8	9785	6	AX427938	AX427938 Sequence
15	2018.6	81.8	9189	6	AX427921	AX427921 Sequence
16	2017	81.7	9167	6	AX427933	AX427933 Sequence
17	2017	81.7	9170	6	AX427928	AX427928 Sequence
18	2017	81.7	9782	6	AX427935	AX427935 Sequence
19	2017	81.7	9783	6	AX427934	AX427934 Sequence
20	2017	81.7	9792	6	AX427932	AX427932 Sequence
21	2015.6	81.6	4353	12	AF287353	AF287353 Synthetic
22	1999	81.0	3015	6	AX455946	AX455946 Sequence
23	1998.4	80.9	2300	6	BD263705	BD263705 Improved
24	1998.4	80.9	2300	6	AR373388	AR373388 Sequence
25	1982.4	80.3	3009	6	AX455987	AX455987 Sequence
26	1969.4	79.8	4319	6	BD263639	BD263639 Improved
27	1969.4	79.8	4319	6	AR373322	AR373322 Sequence
28	1958.8	79.3	4352	12	AF287352	AF287352 Synthetic
29	1907	77.2	2577	6	AX457088	AX457088 Sequence
30	1905.8	77.2	4343	6	AX306428	AX306428 Sequence
31	1901.8	77.0	4341	6	AX306429	AX306429 Sequence
32	1892.2	76.6	1977	6	AX455954	AX455954 Sequence
33	1879.2	76.1	7897	6	AX427923	AX427923 Sequence
34	1858	75.3	2305	6	BD263702	BD263702 Improved
35	1858	75.3	2305	6	AR373385	AR373385 Sequence
36	1837	74.4	1965	6	AX455952	AX455952 Sequence
37	1831.2	74.2	2299	6	BD263703	BD263703 Improved
38	1831.2	74.2	2299	6	AR373386	AR373386 Sequence
39	1825.6	73.9	3024	12	AF287355	AF287355 Synthetic
40	1807.6	73.2	9407	6	AX427937	AX427937 Sequence
41	1678.6	68.0	4307	6	AR302563	AR302563 Sequence
42	1678.6	68.0	4307	6	AX019132	AX019132 Sequence
43	1678.6	68.0	4307	6	BD130468	BD130468 Antiviral
44	1659.4	67.2	4307	6	BD268901	BD268901 Anti-vira
45	1659.4	67.2	4307	6	AX035453	AX035453 Sequence

ALIGNMENTS

RESULT 1
AX455914
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 30 from Patent WO0204493.
ACCESSION AX455914
VERSION AX455914.1 GI:21714899
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;

CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)

FEATURES Location/Qualifiers
source 1. .2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975(+)"

ORIGIN

Query Match 100.0%; Score 2469; DB 6; Length 2469;
Best Local Similarity 100.0%; Pred. No. 4.6e-250;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
```

```
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2461 GGTGAATTC 2469

Db       2461 GGTGAATTC 2469
```

```
RESULT 2
AX455915
LOCUS       AX455915                2463 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 31 from Patent WO0204493.
ACCESSION   AX455915
VERSION     AX455915.1  GI:21714900
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 31 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1. .2463
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YM"
ORIGIN
```

```
  Query Match             98.9%;  Score 2442.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 3e-247;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300

Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
```

```
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
```

```
Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2455 GGTGAATTC 2463
```

RESULT 3
AX455916

LOCUS	AX455916	2457 bp	DNA	linear	PAT 06-JUL-2002

DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS .
SOURCE synthetic construct
  ORGANISM synthetic construct
  artificial sequences.
REFERENCE 1
  AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
  JOURNAL Patent: WO 0204493-A 32 17-JAN-2002;
  CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES Location/Qualifiers
     source 1. .2457
  /organism="synthetic construct"
  /mol_type="unassigned DNA"
  /db_xref="taxon:32630"
  /note="PR975YMWM"
ORIGIN

Query Match 97.8%; Score 2415.4; DB 6; Length 2457;
Best Local Similarity 99.3%; Pred. No. 1.9e-244;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Qy        121 CACATCGCCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
```

```
Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db      1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy      1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db      1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy      1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db      1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy      1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db      1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy      1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db      1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy      1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db      1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy      1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db      1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy      1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db      1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy      1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db      1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy      1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db      1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy      1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db      1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy      1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db      1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy      1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db      1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy      1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db      1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy      2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Db      2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088

Qy      2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db      2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy      2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db      2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy      2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db      2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy      2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db      2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy      2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Db      2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy      2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Db      2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy      2461 GGTGAATTC 2469
Db      2449 GGTGAATTC 2457
```

RESULT 4
BD263706
LOCUS BD263706 2312 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION BD263706
VERSION BD263706.1 GI:33073474
KEYWORDS JP 2002533124-A/73.

```
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH   Key
                     Location/Qualifiers
            FT   source          1. .2312
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             83.1%;  Score 2052;  DB 6;  Length 2312;
  Best Local Similarity   93.6%;  Pred. No. 2.3e-206;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | |||||  || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              ||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||| | ||||     |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      |||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | ||||||||| |||||||| |||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| |||||||||||||    | |||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | |||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
```

```
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              |||||||||||||||||  |||||| ||||||||||||||||||||||||||||||  ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
              |||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
              |||||||||||||||||||||||||| |||||||||||||||||||| ||||||||||||
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
              |||||||| |||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
              |||||||  | |||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
              |||||  |||||||||||||| |||||||||||||||||||||||||||||||||||||
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
              |||||||||||||||||||| |||||| ||||||||||||||||||||||||||||||||
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
              |||||||||||||||| |  || |||||||||||||||||||||||| ||||||||
```

```
Db    1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy    1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db    1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy    1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db    1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy    1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db    1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy    2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db    1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy    2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db    1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy    2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db    1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy    2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db    2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy    2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db    2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy    2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db    2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy    2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db    2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy    2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db    2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 5
AR373389
LOCUS       AR373389                2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             83.1%;  Score 2052;  DB 6;  Length 2312;
  Best Local Similarity   93.6%;  Pred. No. 2.3e-206;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;
```

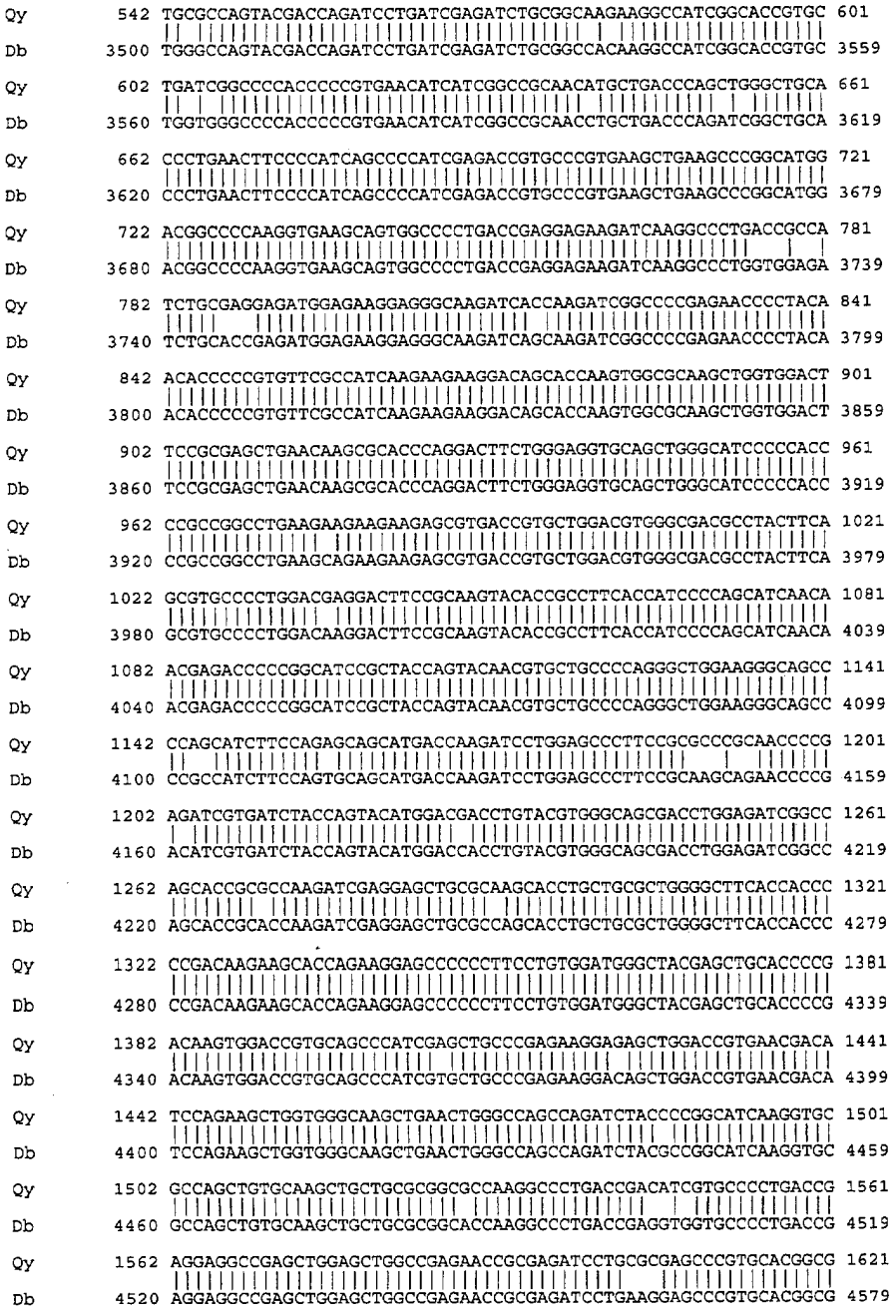
```
Qy     170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db       1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy     230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db      61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy     290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db     121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy     344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db     181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy     398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db     241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy     458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db     301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy     518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db     361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy     578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db     421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy     638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db     481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy     698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db     541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy     758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db     601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy     818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db     661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy     878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db     721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy     938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db     781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy     998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db     841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy    1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db     901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy    1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db     961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy    1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db    1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy    1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
```

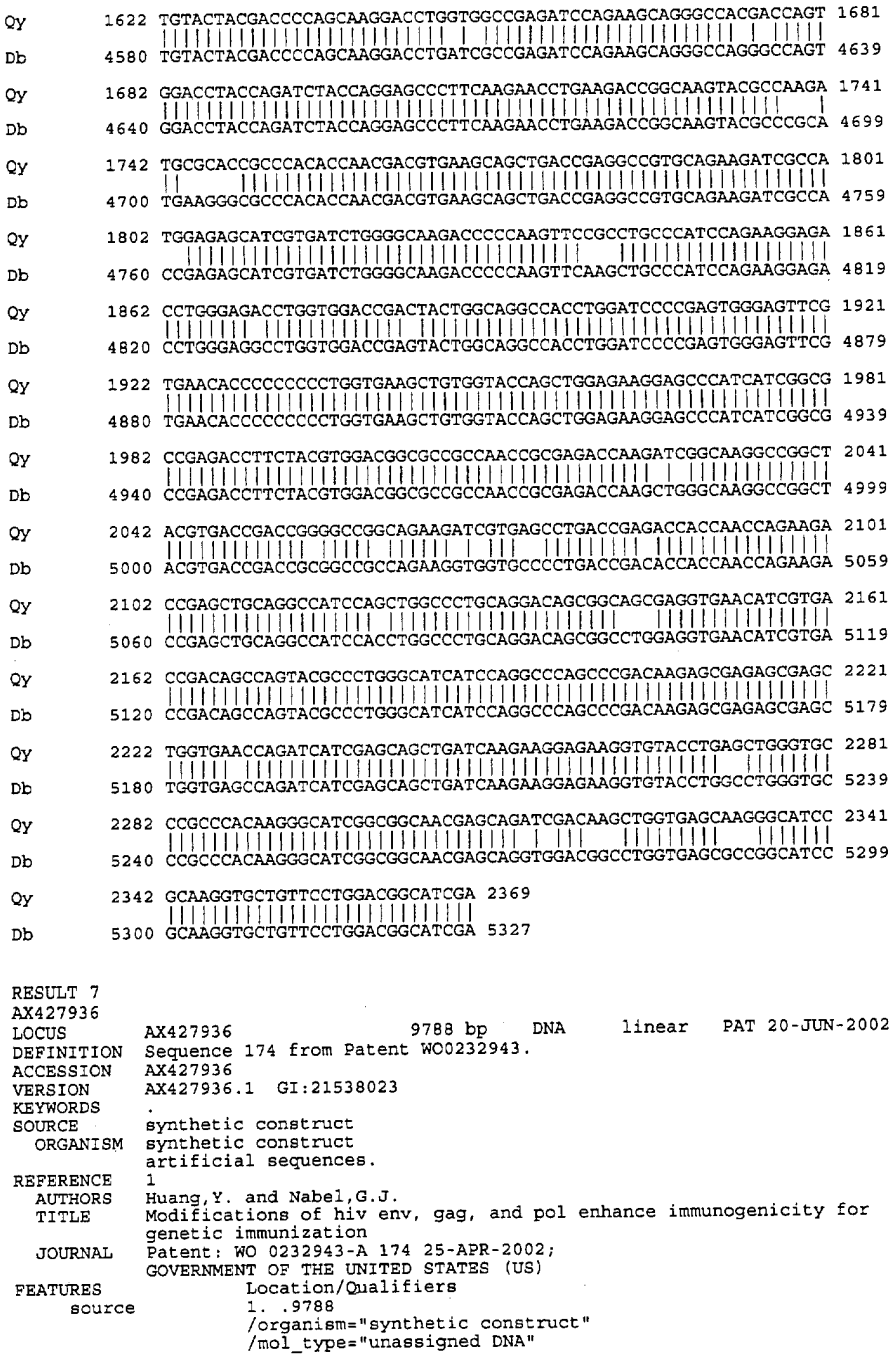
```
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140
Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200
Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260
Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320
Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380
Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440
Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500
Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560
Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620
Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680
Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740
Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800
Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860
Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920
Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980
Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040
Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100
Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160
Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220
Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280
Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

RESULT 6
AX427930
LOCUS AX427930 9166 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 168 from Patent WO0232943.
ACCESSION AX427930
VERSION AX427930.1 GI:21538017
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 168 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES Location/Qualifiers
source 1. .9166
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

Query Match 83.1%; Score 2051.2; DB 6; Length 9166;
Best Local Similarity 92.6%; Pred. No. 2e-206;
Matches 2192; Conservative 0; Mismatches 163; Indels 13; Gaps 3;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080
Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199
Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCCACCAGCCGCGAGCTGC 313
Db       3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259
Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3260 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3319
Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3379
Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3380 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3439
Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3440 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3499
```



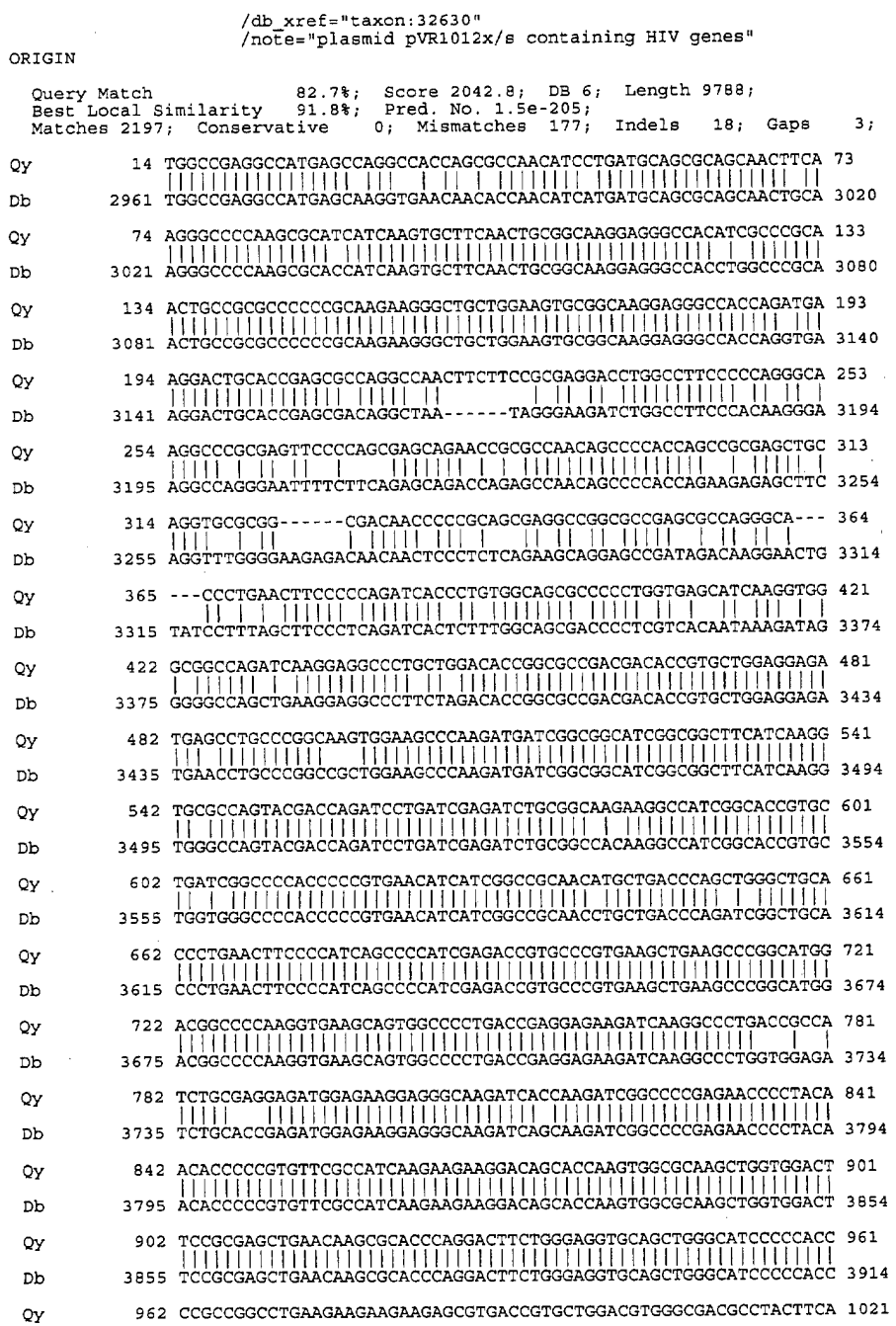


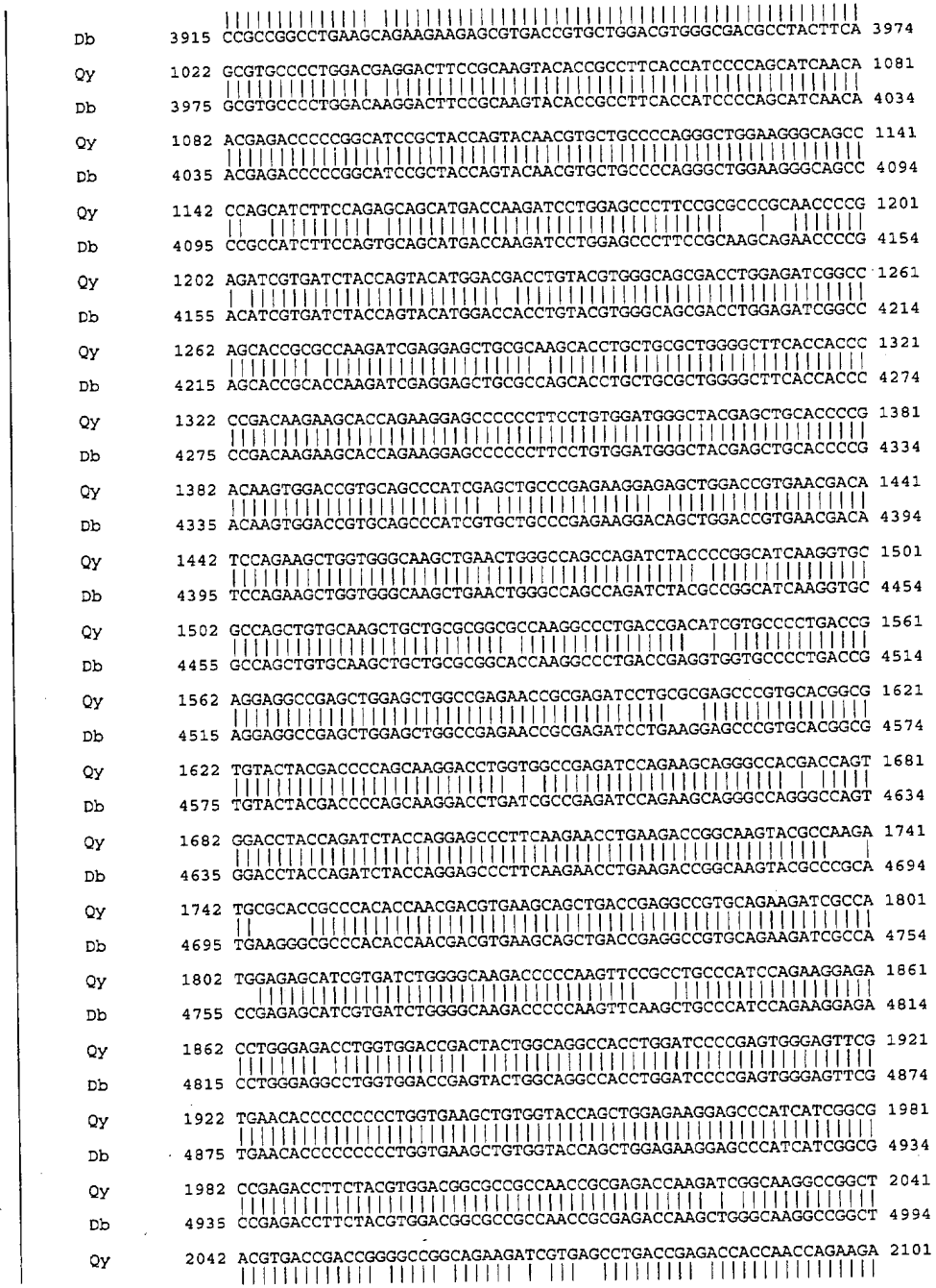
```
RESULT 7
AX427936
LOCUS       AX427936               9788 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
```

/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 82.7%; Score 2042.8; DB 6; Length 9788;
Best Local Similarity 91.8%; Pred. No. 1.5e-205;
Matches 2197; Conservative 0; Mismatches 177; Indels 18; Gaps 3;





```
Db       4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054

Qy       2102 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2161

Db       5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114

Qy       2162 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2221

Db       5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

Qy       2222 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2281

Db       5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234

Qy       2282 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2341

Db       5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy       2342 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393

Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346


RESULT 8
AX427931
LOCUS       AX427931                9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.2%;  Score 2028.8;  DB 6;  Length 9169;
  Best Local Similarity   91.7%;  Pred. No. 4.4e-204;
  Matches 2194;  Conservative    0;  Mismatches  182;  Indels   16;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364

Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421

Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481

Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541

Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601

Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661

Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721

Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781

Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841

Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901

Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961

Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021

Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081

Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141

Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201

Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162

Qy       1202 AGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1261

Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1262 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1321

Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1322 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1381

Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1441

Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402
```

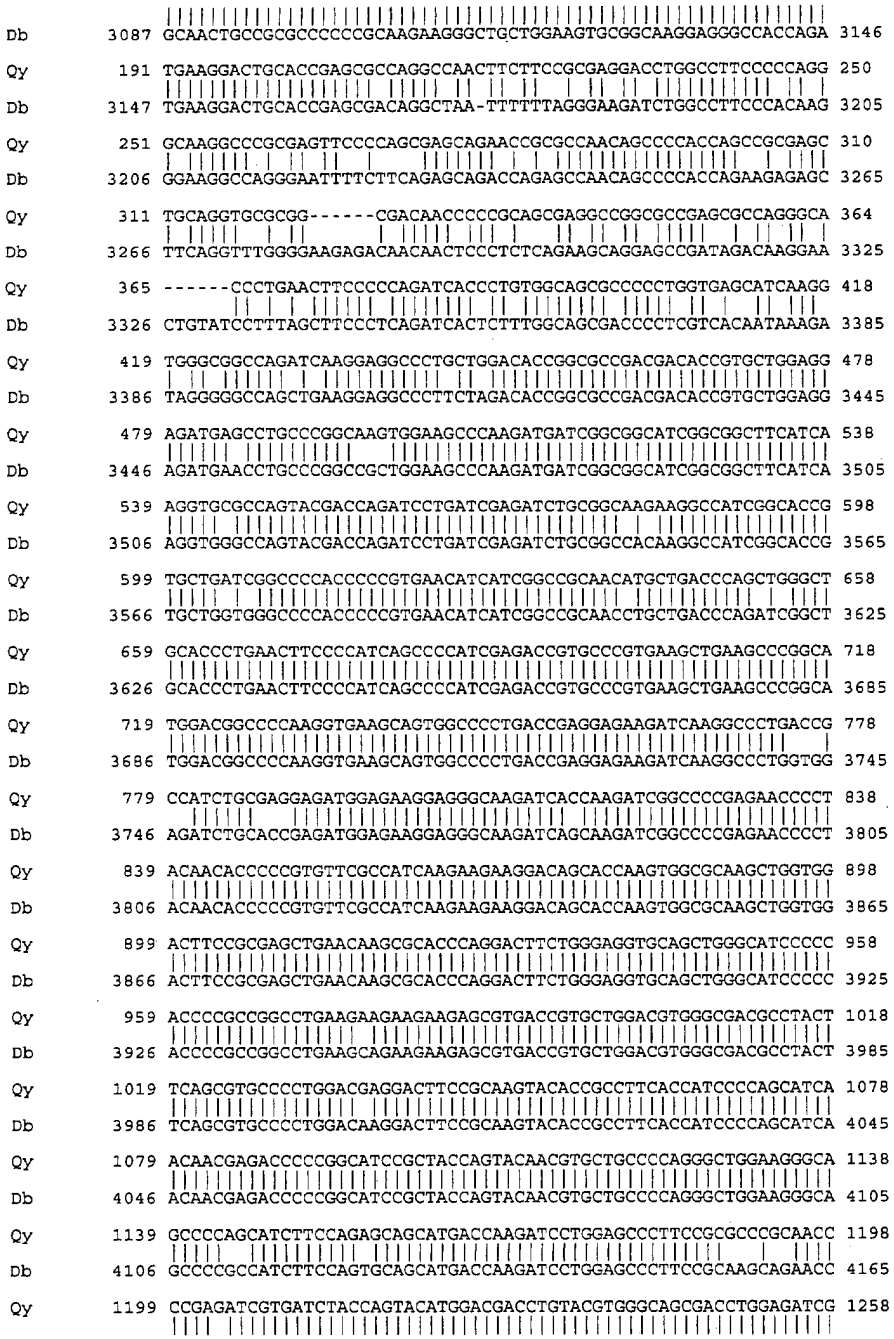
```
Qy       1442 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1501
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1502 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1561
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1562 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1621
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1622 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1681
Db       4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy       1682 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1741
Db       4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy       1742 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1801
Db       4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy       1802 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1861
Db       4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy       1862 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1921
Db       4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy       1922 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1981
Db       4883 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy       1982 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2041
Db       4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy       2042 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2101
Db       5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy       2102 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2161
Db       5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy       2162 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2221
Db       5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy       2222 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2281
Db       5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy       2282 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2341
Db       5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy       2342 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354
```

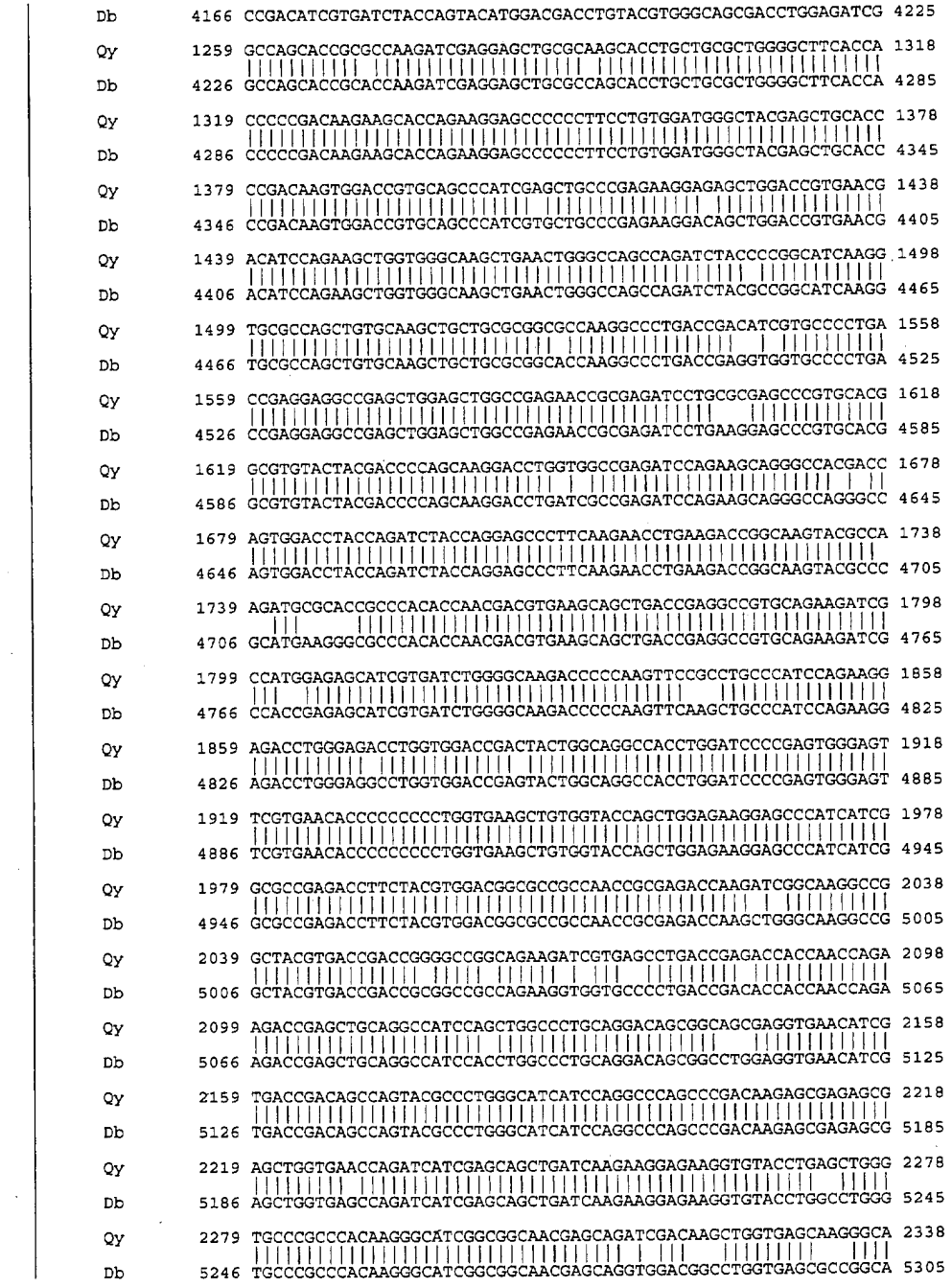
```
RESULT 9
AX427925
LOCUS       AX427925                9194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 163 from Patent WO0232943.
ACCESSION   AX427925
VERSION     AX427925.1  GI:21538012
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 163 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.1%;  Score 2027;  DB 6;  Length 9194;
  Best Local Similarity   91.6%;  Pred. No. 6.8e-204;
  Matches 2194;  Conservative    0;  Mismatches  185;  Indels   16;  Gaps    4;
```

```
Qy         14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
Db       2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db       3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db       3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db       3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy        599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db       3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy        659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db       3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy        719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db       3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745
```

```
Qy        779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db       3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy        839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db       3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy        899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db       3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy        959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db       3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db       3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db       4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db       4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy       1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
Db       4166 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy       1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db       4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy       1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db       4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy       1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db       4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405

Qy       1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db       4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy       1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db       4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy       1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db       4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy       1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db       4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645

Qy       1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db       4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy       1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db       4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy       1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db       4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825

Qy       1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db       4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885

Qy       1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db       4886 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945

Qy       1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db       4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005

Qy       2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db       5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065

Qy       2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db       5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db       5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db       5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db       5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

```
RESULT 10
AX427927
LOCUS       AX427927               12411 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 165 from Patent WO0232943.
ACCESSION   AX427927
VERSION     AX427927.1  GI:21538014
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 165 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .12411
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.1%;  Score 2027;  DB 6;  Length 12411;
  Best Local Similarity   91.6%;  Pred. No. 6.3e-204;
  Matches 2194;  Conservative    0;  Mismatches  185;  Indels   16;  Gaps    4;

Qy          14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
Db        2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy          71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db        3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy         131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
```


Db 3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146
Qy 191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db 3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205
Qy 251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db 3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265
Qy 311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db 3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325
Qy 365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db 3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385
Qy 419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db 3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445
Qy 479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db 3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505
Qy 539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db 3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565
Qy 599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db 3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625
Qy 659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db 3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685
Qy 719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db 3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745
Qy 779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db 3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805
Qy 839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db 3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865
Qy 899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db 3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925
Qy 959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db 3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985
Qy 1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db 3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045
Qy 1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db 4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105
Qy 1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db 4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165
Qy 1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258


Db 4166 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 4225
Qy 1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db 4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285
Qy 1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db 4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345
Qy 1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db 4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405
Qy 1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db 4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465
Qy 1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db 4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525
Qy 1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db 4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585
Qy 1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db 4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645
Qy 1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db 4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705
Qy 1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db 4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765
Qy 1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db 4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825
Qy 1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db 4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885
Qy 1919 TCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db 4886 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945
Qy 1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db 4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005
Qy 2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db 5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065
Qy 2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db 5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125
Qy 2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db 5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185
Qy 2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db 5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245
Qy 2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db 5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

```
Qy     2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db     5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

```
RESULT 11
AX427926
LOCUS       AX427926                9194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 164 from Patent WO0232943.
ACCESSION   AX427926
VERSION     AX427926.1  GI:21538013
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 164 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match            82.0%;  Score 2025.4;  DB 6;  Length 9194;
  Best Local Similarity  91.6%;  Pred. No. 1e-203;
  Matches 2193;  Conservative    0;  Mismatches  186;  Indels   16;  Gaps    4;

Qy       14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
Db     2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy       71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db     3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy      131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db     3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy      191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db     3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy      251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db     3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy      311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db     3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy      365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db     3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy      419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db     3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy      479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db     3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy      539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db     3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy      599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db     3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy      659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db     3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy      719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db     3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745

Qy      779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db     3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy      839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db     3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy      899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db     3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy      959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db     3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy     1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db     3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy     1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db     4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy     1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db     4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy     1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
Db     4166 CCGACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy     1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db     4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy     1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db     4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy     1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db     4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405

Qy     1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db     4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy     1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db     4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy     1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db     4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy     1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db     4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645
```

```
Qy       1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy       1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
               |||     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy       1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
               |||  ||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db       4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825

Qy       1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
               |||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||
Db       4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885

Qy       1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4886 TCGTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945

Qy       1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
               |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db       4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005

Qy       2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
               |||||||||||||||| ||||| |||||| | ||| |||||||| ||||||||||||||
Db       5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065

Qy       2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
               |||||||| |||||||||||||||||||||||||||||||||||||||||||  |||||
Db       5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
               |||||||||||||||||||||||||||||||| | |||  |||||||||  ||||
Db       5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
               |||||||||||||||||||||||||||||||    | |   | |    || || |
Db       5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

RESULT 12
BD263704
LOCUS BD263704 2306 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION BD263704
VERSION BD263704.1 GI:33073472
KEYWORDS JP 2002533124-A/71.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 2306)
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: JP 2002533124-A 71 08-OCT-2002;
CHIRON CORP
COMMENT OS Artificial Sequence
PN JP 2002533124-A/71
PD 08-OCT-2002
PF 30-DEC-1999 JP 2000591193
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
KARIN HARTOG,
PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
C12N5/10,
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH Key
Location/Qualifiers
FT source 1. .2306
FT /organism='Artificial Sequence'.
FEATURES Location/Qualifiers
source 1. .2306
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN

Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 1.5e-203;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

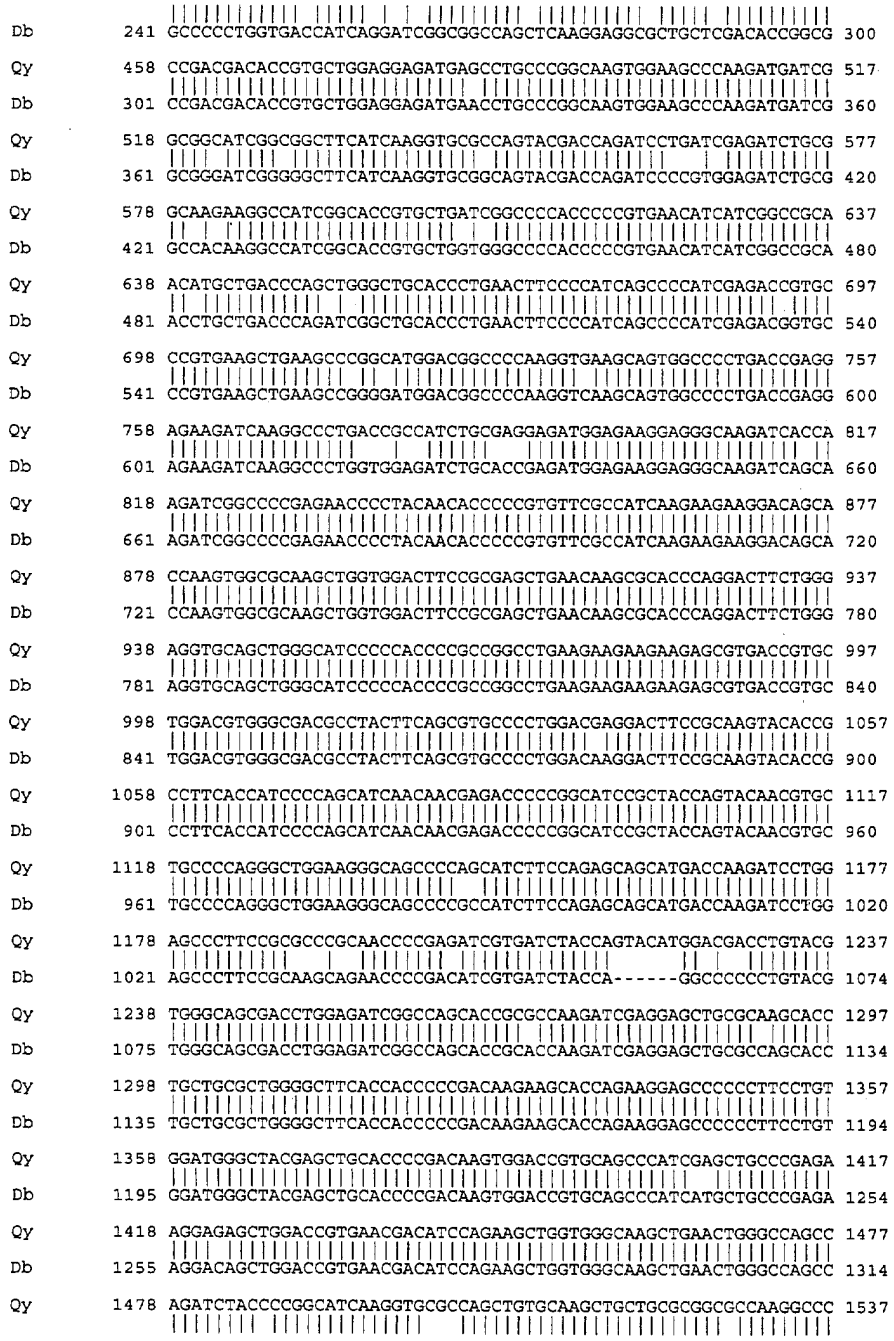
```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              ||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||||| | |||      ||||| || |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      |||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||| ||||| | | |||||||| |||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| |||||||||||||||  | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||||| | |||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | |||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||| || |||||||||||||| |||||||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              |||||||||||||||  | ||||||  |||||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
```

```
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840
Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900
Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960
Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074
Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134
Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194
Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254
Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314
Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374
Qy       1538 TGACCGACATCGTGCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434
Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494
Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554
Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614
Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674
Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734
Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794
Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854
Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914
Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974
Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034
Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094
Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154
Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214
Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274
Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306
```

RESULT 13
AR373387
LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 82 from patent US 6602705.
ACCESSION AR373387
VERSION AR373387.1 GI:40075490
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2306)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 82 05-AUG-2003;
FEATURES Location/Qualifiers
source 1. .2306
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 1.5e-203;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

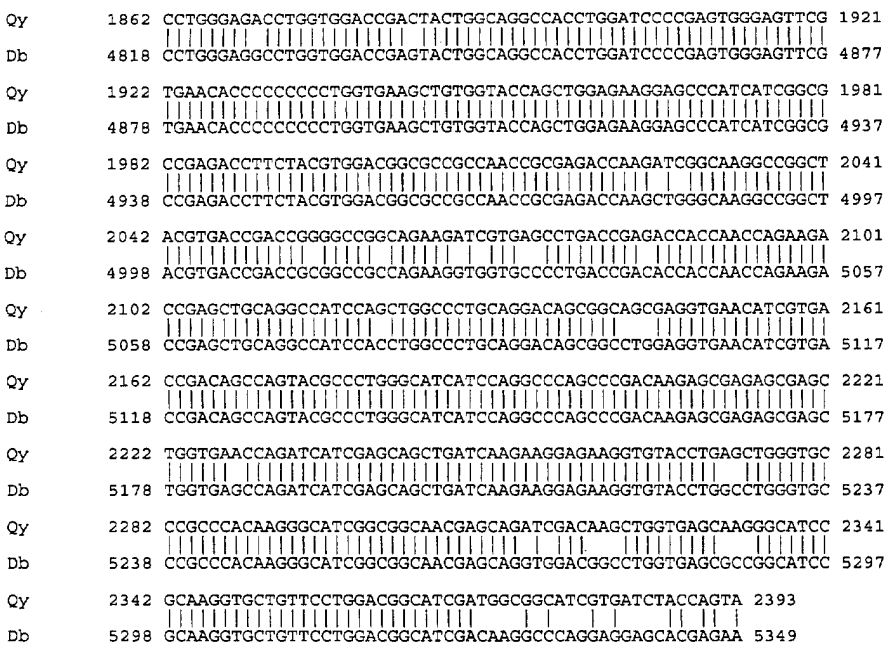
```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60
Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120
Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180
Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
```



```
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300
Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360
Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420
Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480
Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540
Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600
Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660
Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720
Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780
Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840
Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900
Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960
Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074
Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134
Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194
Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254
Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314
Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374
Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434
Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494
Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554
Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614
Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674
Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734
Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794
Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854
Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914
Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974
Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034
Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094
Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154
Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214
Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274
Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306
```

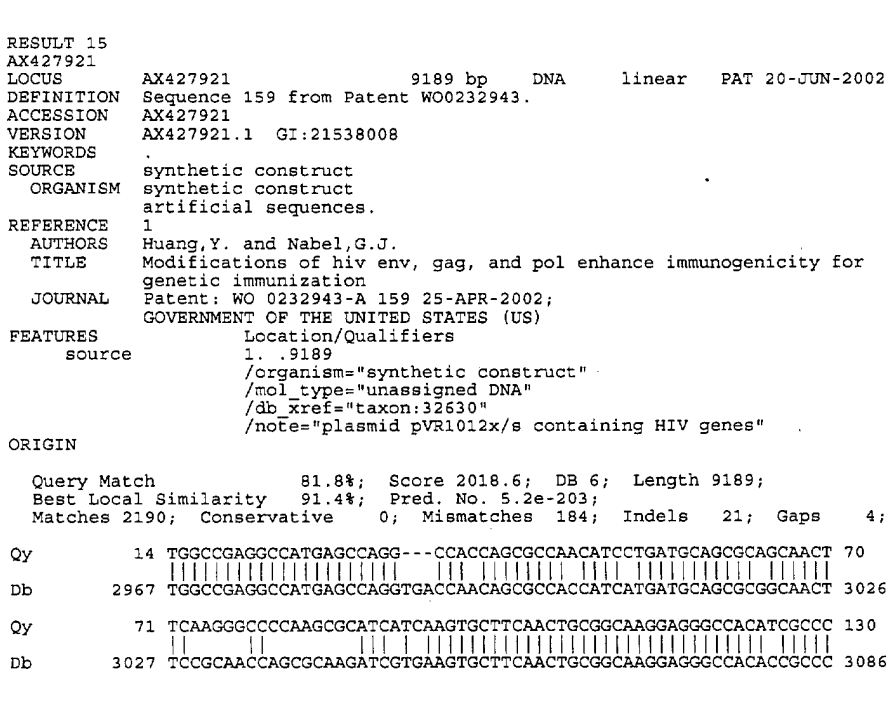
RESULT 14
AX427938
LOCUS AX427938 9785 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 176 from Patent WO0232943.
ACCESSION AX427938
VERSION AX427938.1 GI:21538025

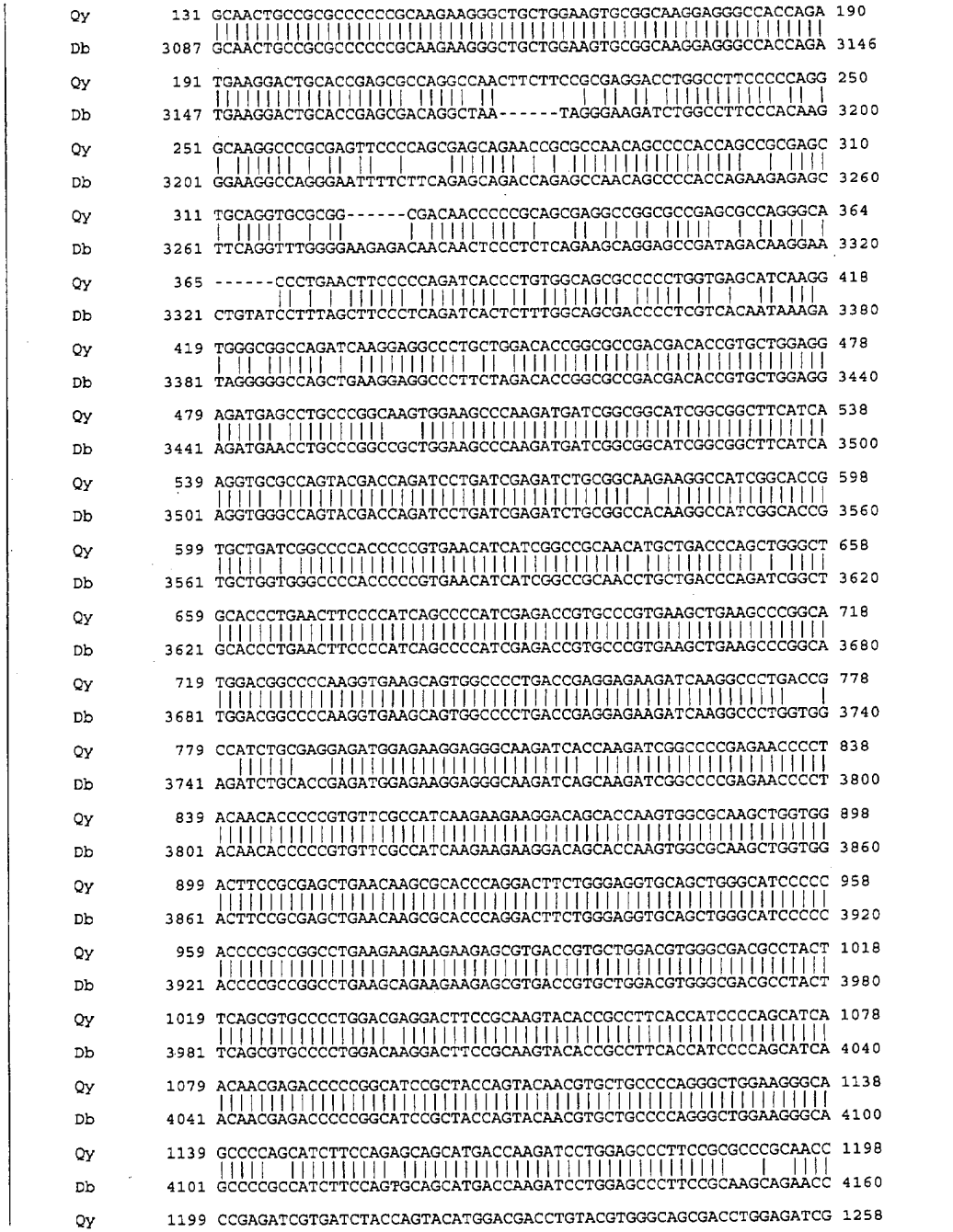
```
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 176 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9785
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             81.8%;  Score 2020.4;  DB 6;  Length 9785;
  Best Local Similarity   91.6%;  Pred. No. 3.3e-203;
  Matches 2190;  Conservative    0;  Mismatches  181;  Indels   21;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3027 AGGGCCAGAAGCG---CATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3144 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3197

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3198 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3257

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3258 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3317

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3318 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3377

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3378 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3437

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3438 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3497

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db       3498 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3557

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db       3558 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3617

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db       3618 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3677

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db       3678 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3737

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db       3738 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3797

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db       3798 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3857

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3858 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3917

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3918 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3977

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3978 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4037

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4038 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4097

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4098 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4157

Qy       1202 AGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1261
Db       4158 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4217

Qy       1262 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1321
Db       4218 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4277

Qy       1322 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1381
Db       4278 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4337

Qy       1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1441
Db       4338 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4397

Qy       1442 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1501
Db       4398 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4457

Qy       1502 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1561
Db       4458 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4517

Qy       1562 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1621
Db       4518 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4577

Qy       1622 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1681
Db       4578 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4637

Qy       1682 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1741
Db       4638 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4697

Qy       1742 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1801
Db       4698 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4757

Qy       1802 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1861
Db       4758 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4817
```



RESULT 15
AX427921
LOCUS AX427921 9189 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 159 from Patent WO0232943.
ACCESSION AX427921
VERSION AX427921.1 GI:21538008
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 159 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES Location/Qualifiers
source 1. .9189
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

Query Match 81.8%; Score 2018.6; DB 6; Length 9189;
Best Local Similarity 91.4%; Pred. No. 5.2e-203;
Matches 2190; Conservative 0; Mismatches 184; Indels 21; Gaps 4;





```
Db      4161 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 4220
Qy      1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db      4221 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4280
Qy      1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db      4281 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4340
Qy      1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db      4341 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4400
Qy      1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db      4401 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4460
Qy      1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db      4461 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4520
Qy      1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db      4521 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4580
Qy      1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db      4581 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4640
Qy      1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db      4641 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4700
Qy      1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db      4701 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4760
Qy      1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db      4761 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4820
Qy      1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db      4821 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4880
Qy      1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db      4881 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4940
Qy      1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db      4941 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5000
Qy      2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db      5001 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5060
Qy      2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db      5061 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5120
Qy      2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db      5121 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5180
Qy      2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db      5181 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5240
Qy      2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db      5241 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5300
Qy      2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db      5301 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5355
```

Search completed: April 10, 2004, 12:49:04
Job time : 6354.08 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:45:46 ; Search time 622.513 Seconds
(without alignments)
16849.133 Million cell updates/sec

Title: US-09-610-313-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2469	100.0	2469	6	ABL39959	Abl39959 Synthetic
2	2457	99.5	2457	7	ACA03548	Aca03548 Synthetic
3	2457	99.5	2457	9	ADC13266	Adc13266 DNA of HI
4	2442.2	98.9	2463	6	ABL39960	Abl39960 Synthetic
5	2434.6	98.6	2457	7	ACA03547	Aca03547 Synthetic
6	2434.6	98.6	2457	9	ADC13265	Adc13265 DNA of HI
7	2415.4	97.8	2457	6	ABL39961	Abl39961 Synthetic
8	2401.8	97.3	2445	7	ACA03546	Aca03546 Synthetic
9	2401.8	97.3	2445	9	ADC13264	Adc13264 DNA of HI
10	2394.8	97.0	3930	9	ADC13230	Adc13230 DNA of HI
11	2393.2	96.9	3930	9	ADC13231	Adc13231 DNA of HI
12	2393.2	96.9	3930	9	ADC13232	Adc13232 DNA of HI
13	2393.2	96.9	5184	7	ACA03591	Aca03591 Synthetic
14	2393.2	96.9	5184	9	ADC13279	Adc13279 DNA of HI
15	2362.8	95.7	3531	9	ADC13234	Adc13234 DNA of HI
16	2361.2	95.6	3537	9	ADC13236	Adc13236 DNA of HI
17	2360.2	95.6	5145	7	ACA03521	Aca03521 Synthetic
18	2360.2	95.6	5145	9	ADC13233	Adc13233 DNA of HI
19	2350.2	95.2	3538	9	ADC13235	Adc13235 DNA of HI
20	2349.4	95.2	3624	7	ACA03550	Aca03550 Synthetic
21	2349.4	95.2	3624	9	ADC13268	Adc13268 DNA of HI
22	2301.8	93.2	3607	7	ACA03551	Aca03551 Synthetic
23	2301.8	93.2	3607	9	ADC13269	Adc13269 DNA of HI
24	2283.6	92.5	3597	7	ACA03549	Aca03549 Synthetic
25	2283.6	92.5	3597	9	ADC13267	Adc13267 DNA of HI
26	2149.6	87.1	2472	7	ACA03543	Aca03543 Synthetic
27	2149.6	87.1	2472	7	ACC78507	Acc78507 HIV p2Pol
28	2121.2	85.9	2466	7	ACA03542	Aca03542 Synthetic
29	2121.2	85.9	2466	7	ACC78506	Acc78506 HIV p2Pol
30	2094.4	84.8	2460	7	ACA03541	Aca03541 Synthetic
31	2094.4	84.8	2460	7	ACC78505	Acc78505 HIV p2Pol
32	2093.4	84.8	3564	7	ACC78488	Acc78488 HIV GagPo
33	2093.4	84.8	3564	7	ACC78489	Acc78489 HIV GagPo
34	2092.8	84.8	4716	7	ACA03522	Aca03522 Synthetic
35	2092.8	84.8	4716	9	ADC13238	Adc13238 DNA of HI
36	2089	84.6	3999	7	ACC78484	Acc78484 HIV GagCo
37	2087.4	84.5	3999	7	ACC78485	Acc78485 HIV GagCo
38	2087.4	84.5	3999	7	ACC78486	Acc78486 HIV GagCo
39	2087.4	84.5	5283	7	ACA03584	Aca03584 Synthetic
40	2087.4	84.5	5283	7	ACC78529	Acc78529 HIV TatRe
41	2087.2	84.5	4713	7	ACA03592	Aca03592 Synthetic
42	2087.2	84.5	4713	9	ADC13280	Adc13280 DNA of HI
43	2086.8	84.5	3462	9	ADC13237	Adc13237 DNA of HI
44	2061.4	83.5	3735	7	ACA03545	Aca03545 Synthetic
45	2061.4	83.5	3735	7	ACC78509	Acc78509 HIV p2Pol

ALIGNMENTS

RESULT 1
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene

CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 100.0%; Score 2469; DB 6; Length 2469;
Best Local Similarity 100.0%; Pred. No. 5.2e-298;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db 1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy 1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db 1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy 1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db 1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy 1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db 1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy 1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db 1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy 1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db 1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy 1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db 1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy 1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db 1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy 1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db 1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy 1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db 1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy 1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db 1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
```

```
Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2461 GGTGAATTC 2469
```

RESULT 2
ACA03548
ID ACA03548 standard; DNA; 2457 BP.
XX
AC ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 36; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 99.5%; Score 2457; DB 7; Length 2457;
Best Local Similarity 100.0%; Pred. No. 1.6e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
```

```
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840
Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1927 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 3
ADC13266
ID ADC13266 standard; DNA; 2457 BP.
XX
AC ADC13266;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 42; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 99.5%; Score 2457; DB 9; Length 2457;
Best Local Similarity 100.0%; Pred. No. 1.6e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
Db    1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy   67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db   61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy  127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
Db  121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy  187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db  181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy  247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db  241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy  307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db  301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy  367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db  361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy  427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db  421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy  487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db  481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy  547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db  541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy  607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db  601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy  667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db  661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy  727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db  721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy  787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db  781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy  847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db  841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy  907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db  901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy  967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db  961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy 1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db 1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy 1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db 1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy 1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
Db 1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy 1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy 1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
Db 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy 1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy 1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
Db 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy 1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
Db 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy 1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
Db 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy 1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
Db 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy 1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
Db 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy 1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
Db 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy 1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
```

```
Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1927 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 4
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
XX
AC ABL39960;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YM SEQ ID NO:31.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 9; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 98.9%; Score 2442.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 1.1e-294;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
```

```
Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db      1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy      1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db      1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy      1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db      1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy      1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db      1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy      1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db      1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy      1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db      1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy      1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db      1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy      1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db      1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy      1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db      1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy      1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db      1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy      1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db      1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy      1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db      1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy      1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db      1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy      1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db      1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy      2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Db      2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy      2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db      2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy      2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db      2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy      2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db      2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy      2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db      2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy      2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Db      2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy      2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Db      2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy      2461 GGTGAATTC 2469
Db      2455 GGTGAATTC 2463
```

```
RESULT 5
ACA03547
ID   ACA03547 standard; DNA; 2457 BP.
XX
AC   ACA03547;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
```

PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 35; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.6%; Score 2434.6; DB 7; Length 2457;
Best Local Similarity 99.6%; Pred. No. 9.7e-294;
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||||  || | ||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
```

```
Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463
             |||
Db       2455 GGT 2457
```

```
RESULT 6
ADC13265
ID   ADC13265 standard; DNA; 2457 BP.
XX
AC   ADC13265;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 41; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match          98.6%;  Score 2434.6;  DB 9;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 9.7e-294;
  Matches 2453;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;
```

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480
```

```
Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db      1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy      1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db      1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy      1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db      1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy      1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db      1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy      1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db      1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy      1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db      1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
```

```
Qy      1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db      1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy      1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db      1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy      1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db      1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy      1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db      1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy      1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db      1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy      1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db      1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy      1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db      1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy      1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db      1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy      2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Db      2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy      2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db      2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy      2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db      2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy      2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db      2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy      2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db      2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy      2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Db      2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy      2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Db      2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy      2461 GGT 2463
Db      2455 GGT 2457
```

```
RESULT 7
ABL39961
ID   ABL39961 standard; DNA; 2457 BP.
XX
AC   ABL39961;
XX
DT   15-MAY-2002  (first entry)
XX
```

DE Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 10; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 97.8%; Score 2415.4; DB 6; Length 2457;
Best Local Similarity 99.3%; Pred. No. 2.3e-291;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

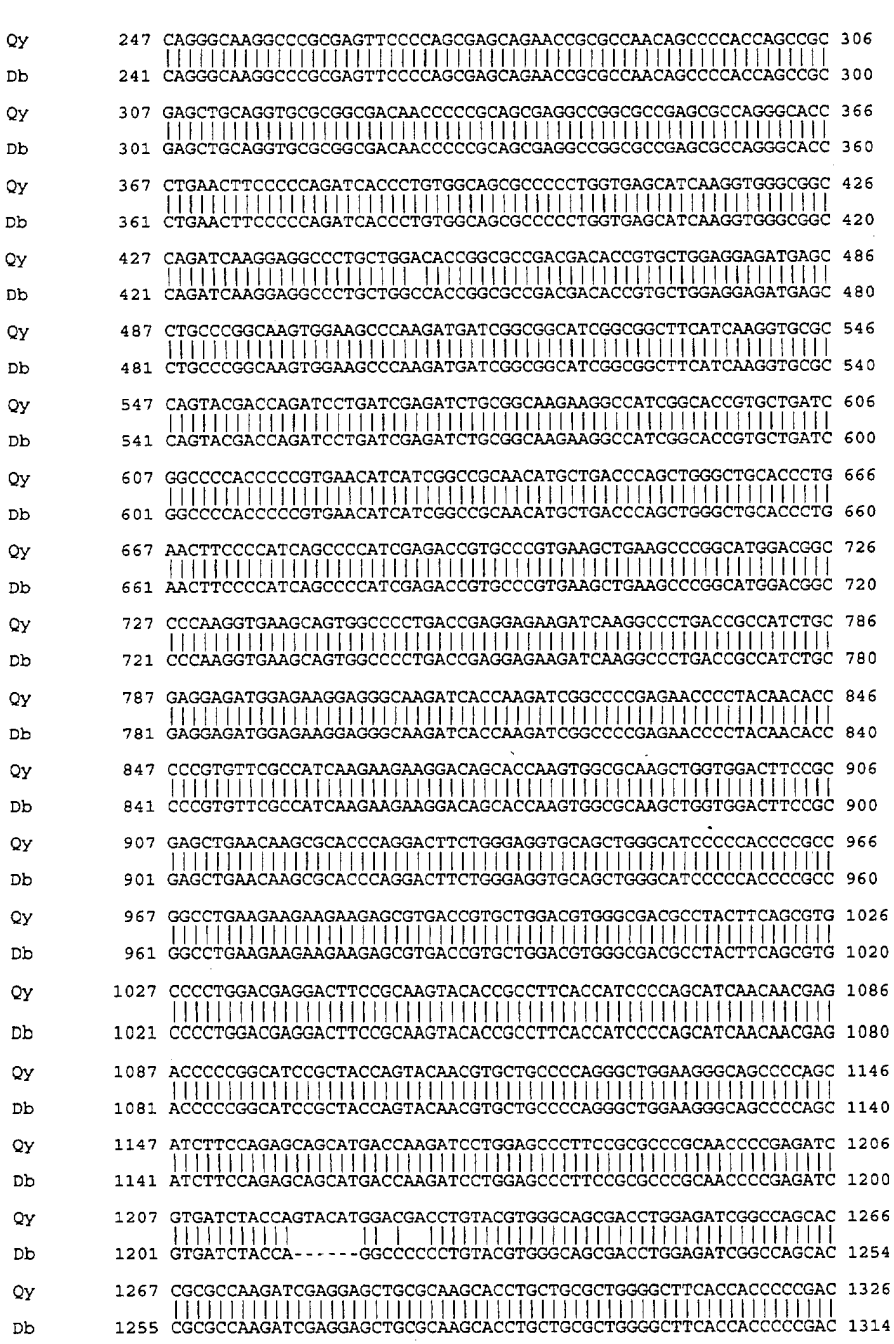
```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||  ||  |||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||  ||      ||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

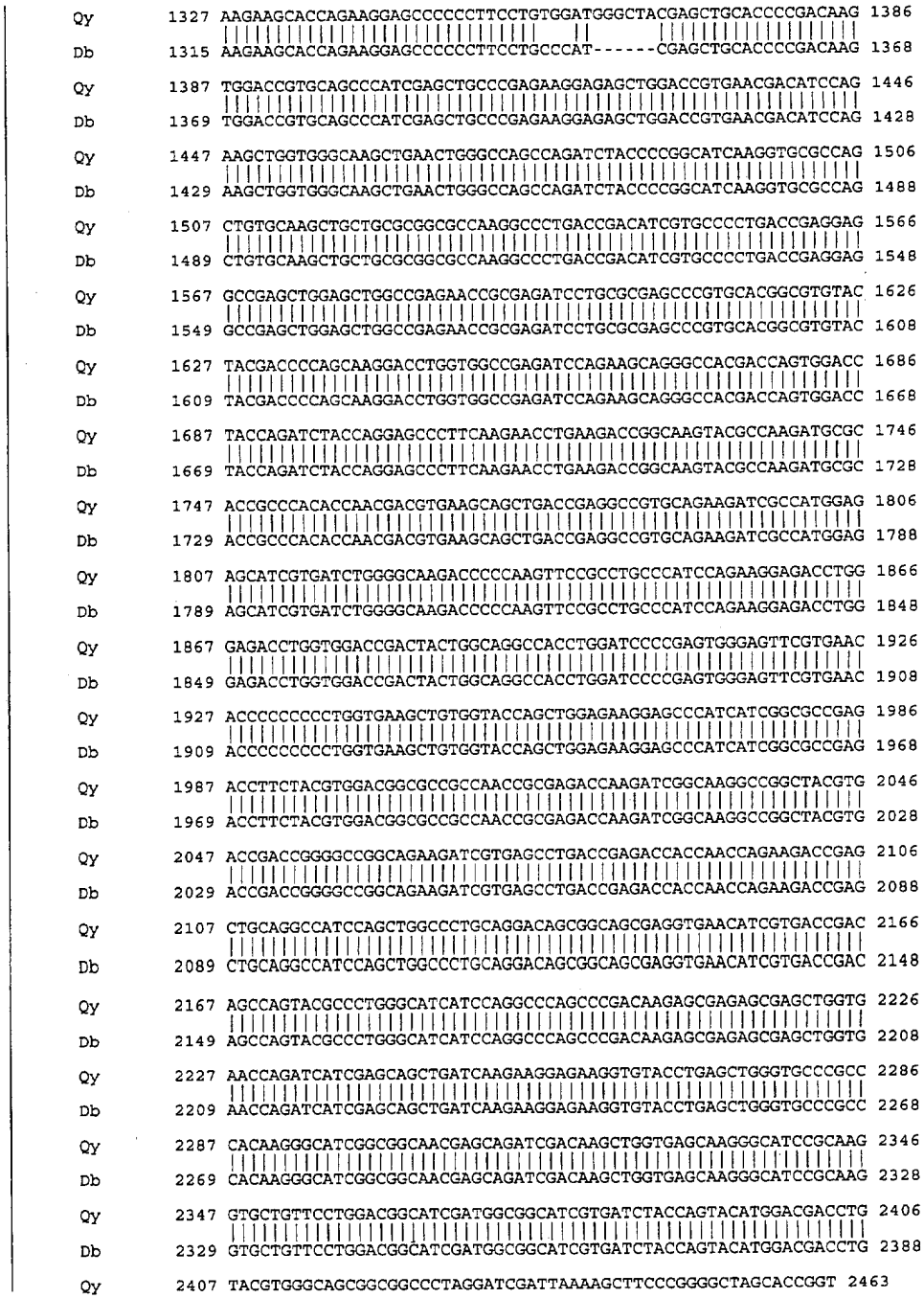
```
Db       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2449 GGTGAATTC 2457
```

RESULT 8
ACA03546

```
ID   ACA03546 standard; DNA; 2445 BP.
XX
AC   ACA03546;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 34; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;
```

```
  Query Match             97.3%;  Score 2401.8;  DB 7;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-289;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240
```





```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

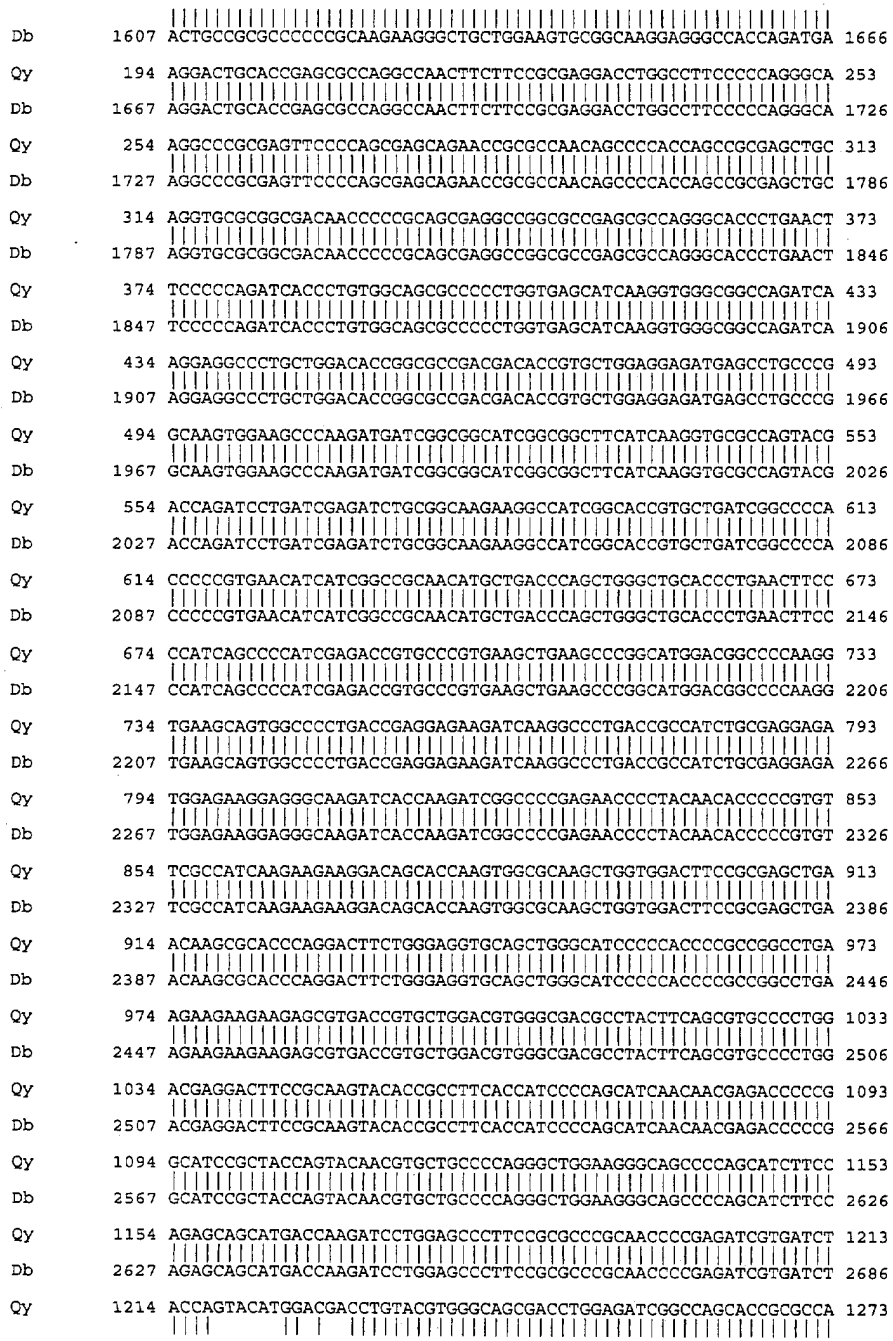
```
RESULT 9
ADC13264
ID   ADC13264 standard; DNA; 2445 BP.
XX
AC   ADC13264;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 40; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match             97.3%;  Score 2401.8;  DB 9;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-289;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              |||||||||||      || |  |||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
```

```
Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386

Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446

Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506

Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566

Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626

Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686

Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746

Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806

Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866

Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926

Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986

Db       1909 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046

Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106

Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166

Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226

Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286

Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346

Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406

Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 10
ADC13230
ID ADC13230 standard; DNA; 3930 BP.
XX
AC ADC13230;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 6; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 97.0%; Score 2394.8; DB 9; Length 3930;
Best Local Similarity 99.2%; Pred. No. 7.8e-289;
Matches 2431; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
```


Db 1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
Qy 194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db 1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db 2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740
Qy 1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db 2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800
Qy 1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db 2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854
Qy 1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db 2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914
Qy 1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db 2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974
Qy 1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db 2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034
Qy 1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db 3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094
Qy 1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db 3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154
Qy 1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db 3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db 3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274
Qy 1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db 3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334
Qy 1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db 3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394
Qy 1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db 3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454
Qy 1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db 3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514
Qy 2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db 3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db 3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634
Qy 2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db 3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694
Qy 2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
Db 3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754
Qy 2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db 3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

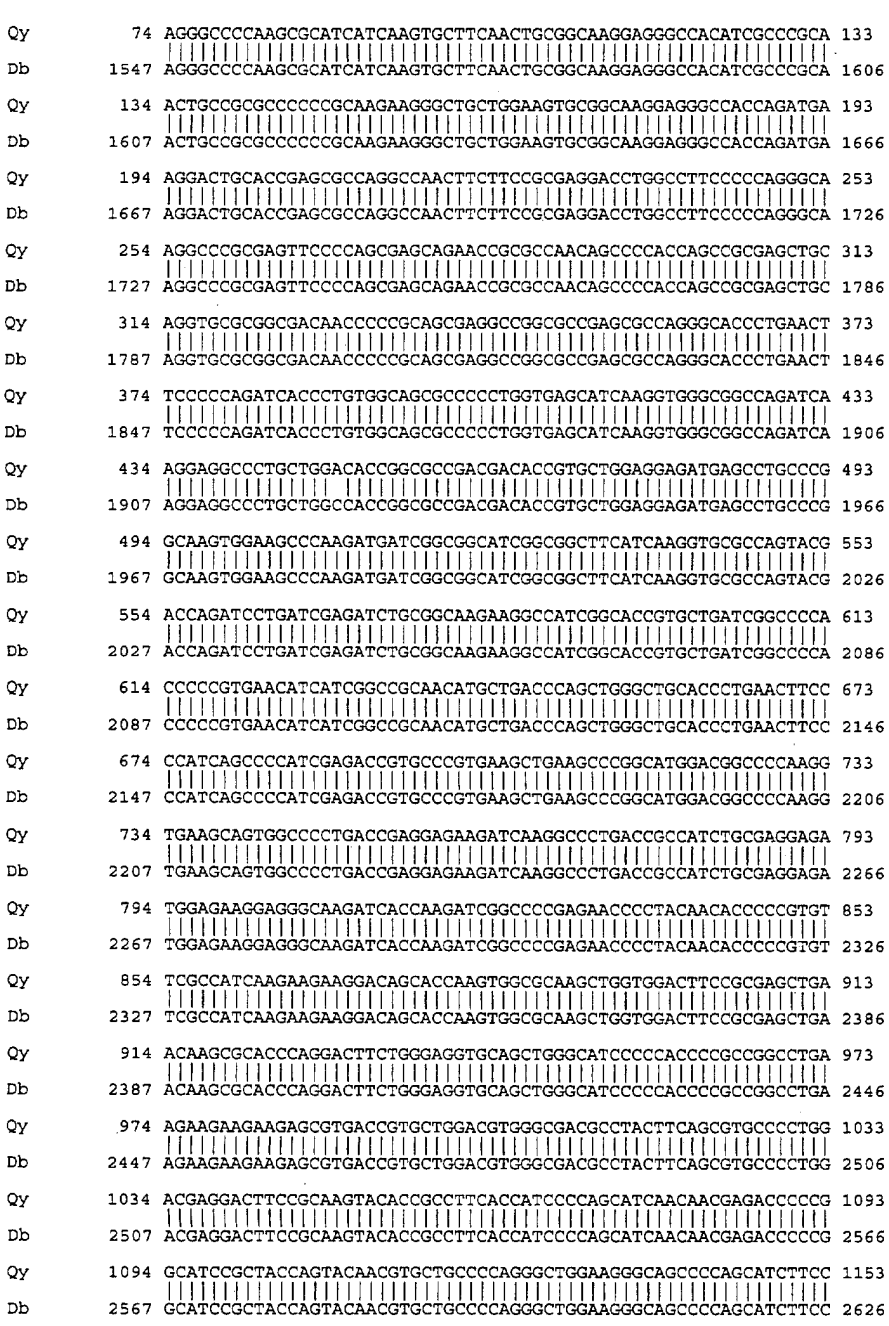
```
Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 11
ADC13231
ID   ADC13231 standard; DNA; 3930 BP.
XX
AC   ADC13231;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutAtt_C SEQ ID NO 10.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 7; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

  Query Match             96.9%;  Score 2393.2;  DB 9;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 1.2e-288;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
```

```
Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
             ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
             ||||||||||||||||||||||  ||       ||||||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 12
ADC13232
ID ADC13232 standard; DNA; 3930 BP.
XX
AC ADC13232;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 8; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match   96.9%; Score 2393.2; DB 9; Length 3930;
  Best Local Similarity 99.2%; Pred. No. 1.2e-288;
  Matches 2430; Conservative   0; Mismatches   8; Indels   12; Gaps   2;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546
```


Qy 74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db 1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606
Qy 134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db 1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
Qy 194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db 1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db 2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740
Qy 1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db 2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800
Qy 1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db 2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854
Qy 1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db 2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914
Qy 1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db 2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974
Qy 1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db 2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034
Qy 1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db 3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094
Qy 1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db 3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154
Qy 1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db 3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db 3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274
Qy 1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db 3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334
Qy 1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db 3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394
Qy 1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db 3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454
Qy 1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db 3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514
Qy 2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db 3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db 3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634
Qy 2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db 3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694
Qy 2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293

```
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353

Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413

Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 13
ACA03591
ID ACA03591 standard; DNA; 5184 BP.
XX
AC ACA03591;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #74.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 79; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match 96.9%; Score 2393.2; DB 7; Length 5184;

Best Local Similarity 99.2%; Pred. No. 1.2e-288;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373

Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433

Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493

Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553

Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613

Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673

Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733

Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793

Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853

Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913

Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973

Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033

Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
```

```
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
             ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 3994

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3995 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 4054

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
             |||||||||||||||||||||    ||        ||||||||||||||||||||||||||||
Db       4055 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 4108

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4109 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 4168

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4169 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 4228

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4229 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 4288

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4289 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 4348

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4349 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 4408

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4409 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 4468

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4469 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 4528

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4529 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 4588

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4589 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 4648

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4649 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 4708

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4709 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 4768

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4769 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 4828

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4829 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 4888

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4889 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 4948

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4949 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 5008

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5009 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 5068

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5069 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 5128

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5129 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 14
ADC13279
ID ADC13279 standard; DNA; 5184 BP.
XX
AC ADC13279;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct TatRevNefgagCpo1Ina_C SEQ ID NO 58.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 55; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX

SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match 96.9%; Score 2393.2; DB 9; Length 5184;
Best Local Similarity 99.2%; Pred. No. 1.2e-288;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

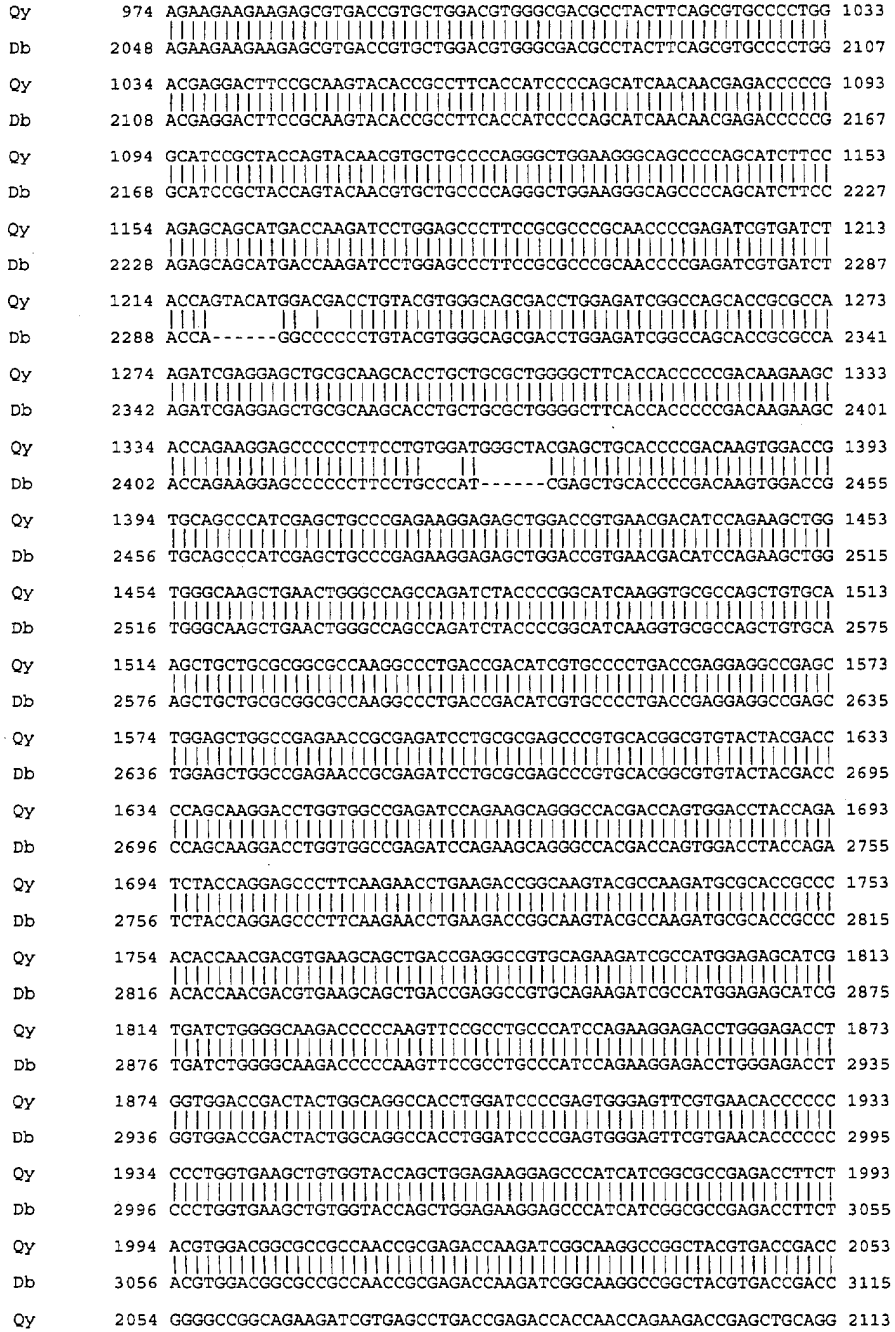
```
Qy      14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy      74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy     134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy     194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy     254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy     314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy     374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy     434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy     494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy     554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy     614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy     674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy     734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy     794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy     854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy     914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy     974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
```

```
Qy    1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy    1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy    1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy    1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
           ||||  ||  |  ||||||||||||||||||||||||||||||||||||||||||||||
Db    3941 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 3994

Qy    1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3995 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 4054

Qy    1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
           ||||||||||||||||||||  ||  ||||||||||||||||||||||||||||||||||
Db    4055 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 4108

Qy    1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4109 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 4168

Qy    1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4169 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 4228

Qy    1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4229 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 4288

Qy    1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4289 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 4348

Qy    1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4349 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 4408

Qy    1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4409 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 4468

Qy    1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4469 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 4528

Qy    1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4529 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 4588

Qy    1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4589 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 4648

Qy    1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4649 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 4708

Qy    1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4709 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 4768

Qy    2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4769 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 4828
```

```
Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4829 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 4888

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4889 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 4948

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4949 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 5008

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5009 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 5068

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5069 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 5128

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5129 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 15
ADC13234
ID ADC13234 standard; DNA; 3531 BP.
XX
AC ADC13234;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagPolmut_C SEQ ID NO 13.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 10; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3531 BP; 802 A; 1210 C; 1096 G; 423 T; 0 U; 0 Other;

```
  Query Match             95.7%;  Score 2362.8;  DB 9;  Length 3531;
  Best Local Similarity   98.4%;  Pred. No. 7.3e-285;
  Matches 2411;  Conservative    0;  Mismatches   27;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| |||||||||||||| || |  |  ||  ||  | || ||| |     |
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047
```



Search completed: April 10, 2004, 07:32:20
Job time : 632.513 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:48:32 ; Search time 118.622 Seconds
(without alignments)
11550.795 Million cell updates/sec

Title: US-09-610-313-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

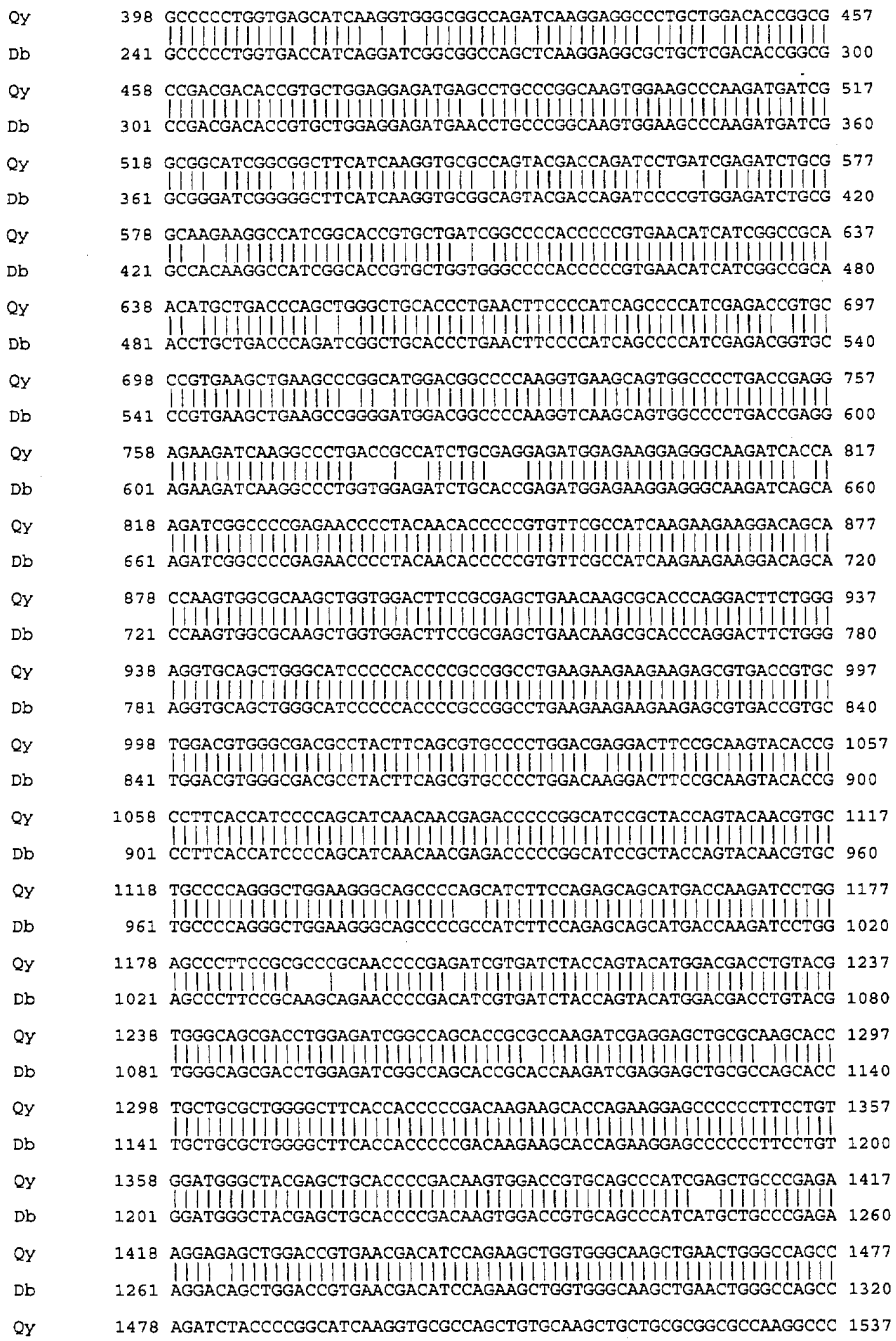
Result No.	Score	% Query Match	Length	DB	ID	Description
1	2052	83.1	2312	4	US-09-475-515-84	Sequence 84, Appl
2	2025.2	82.0	2306	4	US-09-475-515-82	Sequence 82, Appl
3	1998.4	80.9	2300	4	US-09-475-515-83	Sequence 83, Appl
4	1969.4	79.8	4319	4	US-09-475-515-6	Sequence 6, Appli
5	1858	75.3	2305	4	US-09-475-515-80	Sequence 80, Appl
6	1831.2	74.2	2299	4	US-09-475-515-81	Sequence 81, Appl
7	1678.6	68.0	4307	4	US-09-552-950-2	Sequence 2, Appli
8	1651.4	66.9	9772	4	US-09-552-950-5	Sequence 5, Appli
9	1592.8	64.5	8366	4	US-09-872-733A-6	Sequence 6, Appli
10	1557.4	63.1	4338	4	US-09-872-733A-1	Sequence 1, Appli
11	1227.2	49.7	9010	4	US-09-184-418C-8	Sequence 8, Appli
12	1196.2	48.4	8972	4	US-09-184-418C-9	Sequence 9, Appli
13	1190.6	48.2	2467	4	US-09-872-733A-3	Sequence 3, Appli
14	1189	48.2	8959	4	US-09-184-418C-11	Sequence 11, Appl
15	1165.6	47.2	8992	4	US-09-184-418C-4	Sequence 4, Appli
16	1132	45.8	2601	3	US-09-117-217-7	Sequence 7, Appli
17	1132	45.8	2601	3	US-09-117-217-9	Sequence 9, Appli
18	1132	45.8	2601	3	US-09-117-217-11	Sequence 11, Appl
19	1132	45.8	2601	3	US-09-117-217-13	Sequence 13, Appl
20	1132	45.8	2601	4	US-09-735-487-7	Sequence 7, Appli
21	1132	45.8	2601	4	US-09-735-487-9	Sequence 9, Appli
22	1132	45.8	2601	4	US-09-735-487-11	Sequence 11, Appl
23	1132	45.8	2601	4	US-09-735-487-13	Sequence 13, Appl
24	1132	45.8	4307	4	US-09-552-950-1	Sequence 1, Appli
25	1132	45.8	9719	4	US-09-700-304-1	Sequence 1, Appli
26	1128.8	45.7	9050	4	US-09-184-418C-7	Sequence 7, Appli
27	1125.6	45.6	7399	2	US-08-418-848A-9	Sequence 9, Appli
28	1125.6	45.6	9709	2	US-08-188-583-5	Sequence 5, Appli
29	1125.6	45.6	9709	3	US-08-388-353-1	Sequence 1, Appli
30	1125.6	45.6	9709	3	US-08-488-551B-1	Sequence 1, Appli
31	1125.6	45.6	9709	4	US-09-309-572-15	Sequence 15, Appl
32	1125.6	45.6	9709	4	US-09-718-096-15	Sequence 15, Appl
33	1125.6	45.6	12479	4	US-09-318-138-13	Sequence 13, Appl
34	1125.6	45.6	12494	3	US-08-935-312-13	Sequence 13, Appl
35	1125.6	45.6	12494	3	US-08-848-760B-33	Sequence 33, Appl
36	1125.6	45.6	15581	3	US-08-646-538-35	Sequence 35, Appl
37	1125.6	45.6	15581	3	US-09-503-222-35	Sequence 35, Appl
38	1122.8	45.5	3000	4	US-09-184-418C-74	Sequence 74, Appl
39	1121	45.4	8968	4	US-09-184-418C-1	Sequence 1, Appli
40	1116	45.2	4313	4	US-09-475-515-3	Sequence 3, Appli
41	1116	45.2	9737	2	US-08-944-449-7	Sequence 7, Appli
42	1116	45.2	9737	4	US-09-353-362-7	Sequence 7, Appli
43	1108	44.9	8954	4	US-09-184-418C-6	Sequence 6, Appli
44	1101.2	44.6	3017	4	US-09-184-418C-83	Sequence 83, Appl
45	1099	44.5	3011	4	US-09-184-418C-101	Sequence 101, App

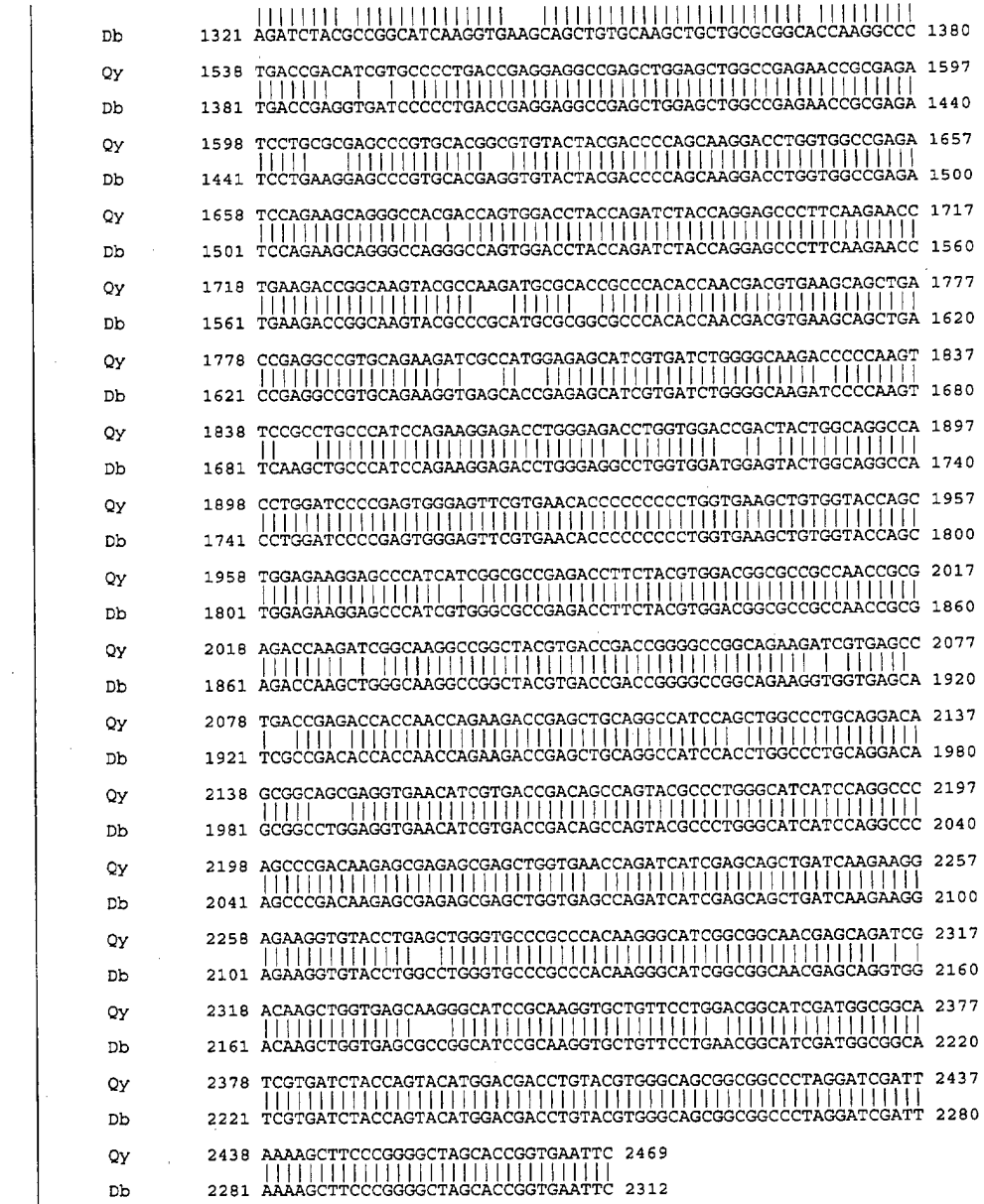
ALIGNMENTS

RESULT 1
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 84
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt(+)
US-09-475-515-84

Query Match 83.1%; Score 2052; DB 4; Length 2312;
Best Local Similarity 93.6%; Pred. No. 0;
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

```
Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
            |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
            ||||||||||||||  ||||||||||||||||||||  |||||||||| ||||||||
Db       61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
            |||||||||| |||||||||||||||| | ||||      |||||  ||  |||||||
Db      121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy      344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
            |||||||||| |||||||||||  ||      |||||||||||||||||||||||||||
Db      181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
```





RESULT 2
US-09-475-515-82
; Sequence 82, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:

```
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 2306
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

  Query Match            82.0%;  Score 2025.2;  DB 4;  Length 2306;
  Best Local Similarity  93.3%;  Pred. No. 0;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
```

```
Db       1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017

Db       1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077

Db       1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137

Db       1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197

Db       1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257

Db       2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317

Db       2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377

Db       2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437

Db       2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469

Db       2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306
```

RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YMWM
US-09-475-515-83

Query Match 80.9%; Score 1998.4; DB 4; Length 2300;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

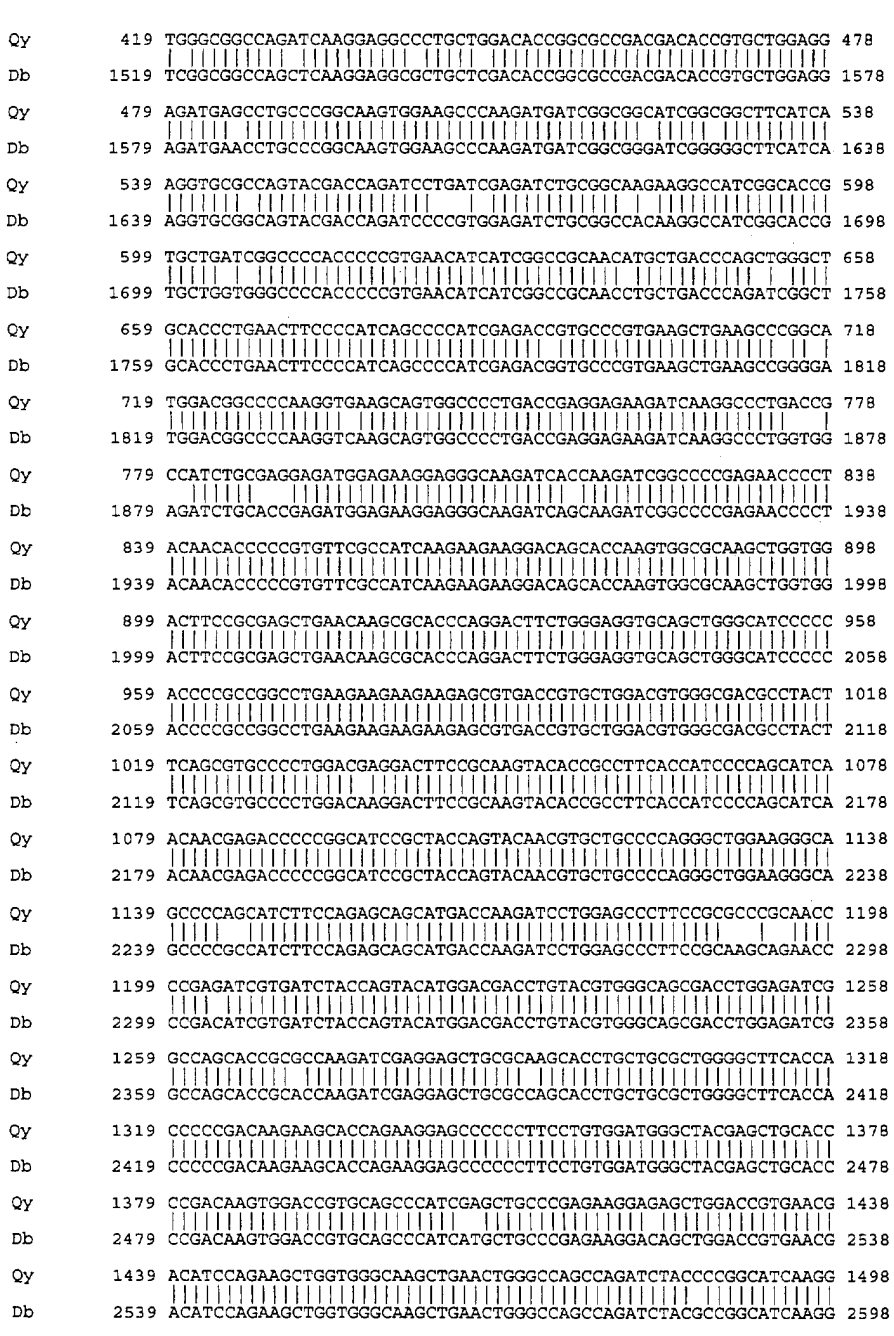
```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074
```

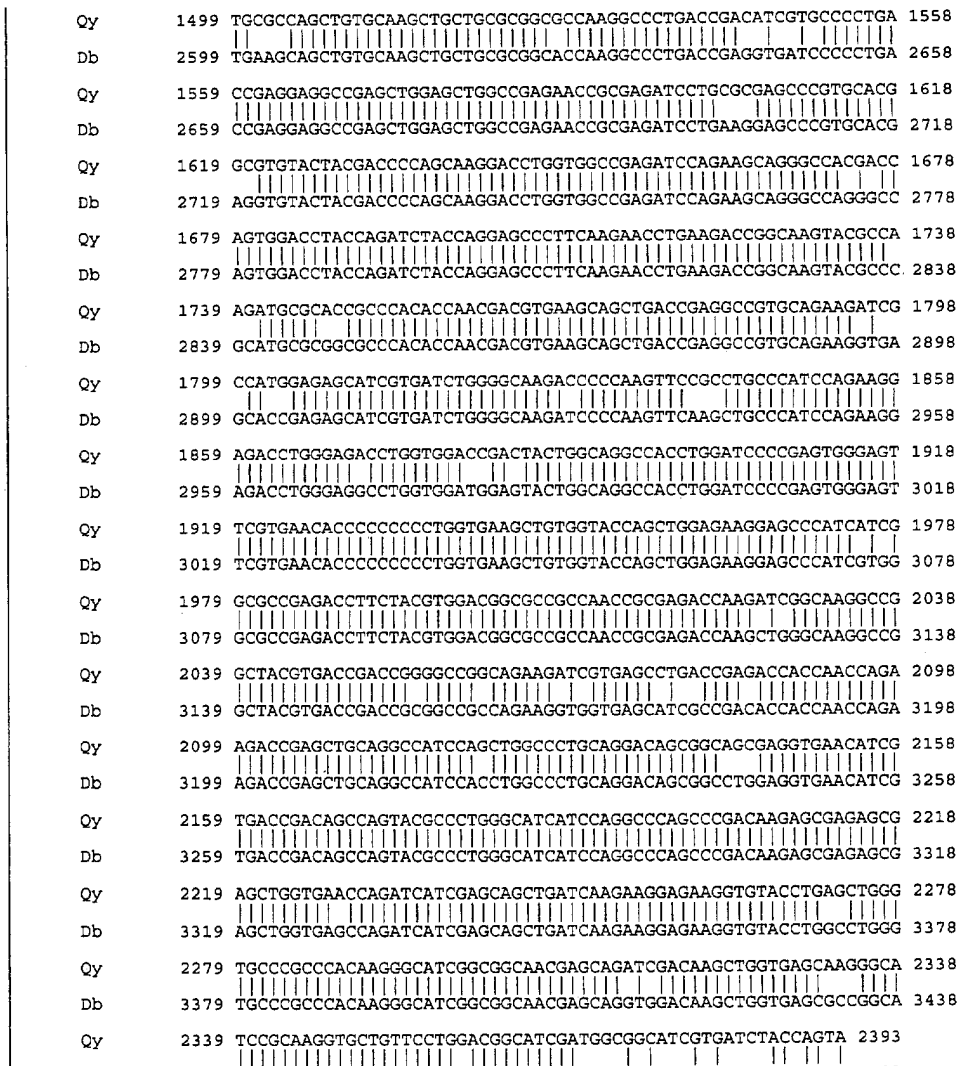
```
Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1194

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1248

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1249 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1308

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1309 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1368

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1369 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1428

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1429 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1488

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1489 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1548

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1549 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1608

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1609 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1668

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1669 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1728

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1729 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1788

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1789 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1848

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1849 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1908

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1909 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1968

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1969 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2028

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2029 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2088

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2089 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2148

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2149 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2208

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2209 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2268

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2269 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2300
```

RESULT 4
US-09-475-515-6
; Sequence 6, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: HIV-Gag-polymerase
US-09-475-515-6

Query Match 79.8%; Score 1969.4; DB 4; Length 4319;
Best Local Similarity 90.1%; Pred. No. 0;
Matches 2158; Conservative 0; Mismatches 221; Indels 16; Gaps 4;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db       1100 TGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCGACCATCATGATGCAGCGCGGCAACT 1159

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       1160 TCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCA 1219

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       1220 GGAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCGAAGGACACCAAA 1279

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       1280 TGAAAGATTGCACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 1338

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       1339 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 1398

Qy        311 TGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       1399 TTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 1458

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       1459 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 1518
```

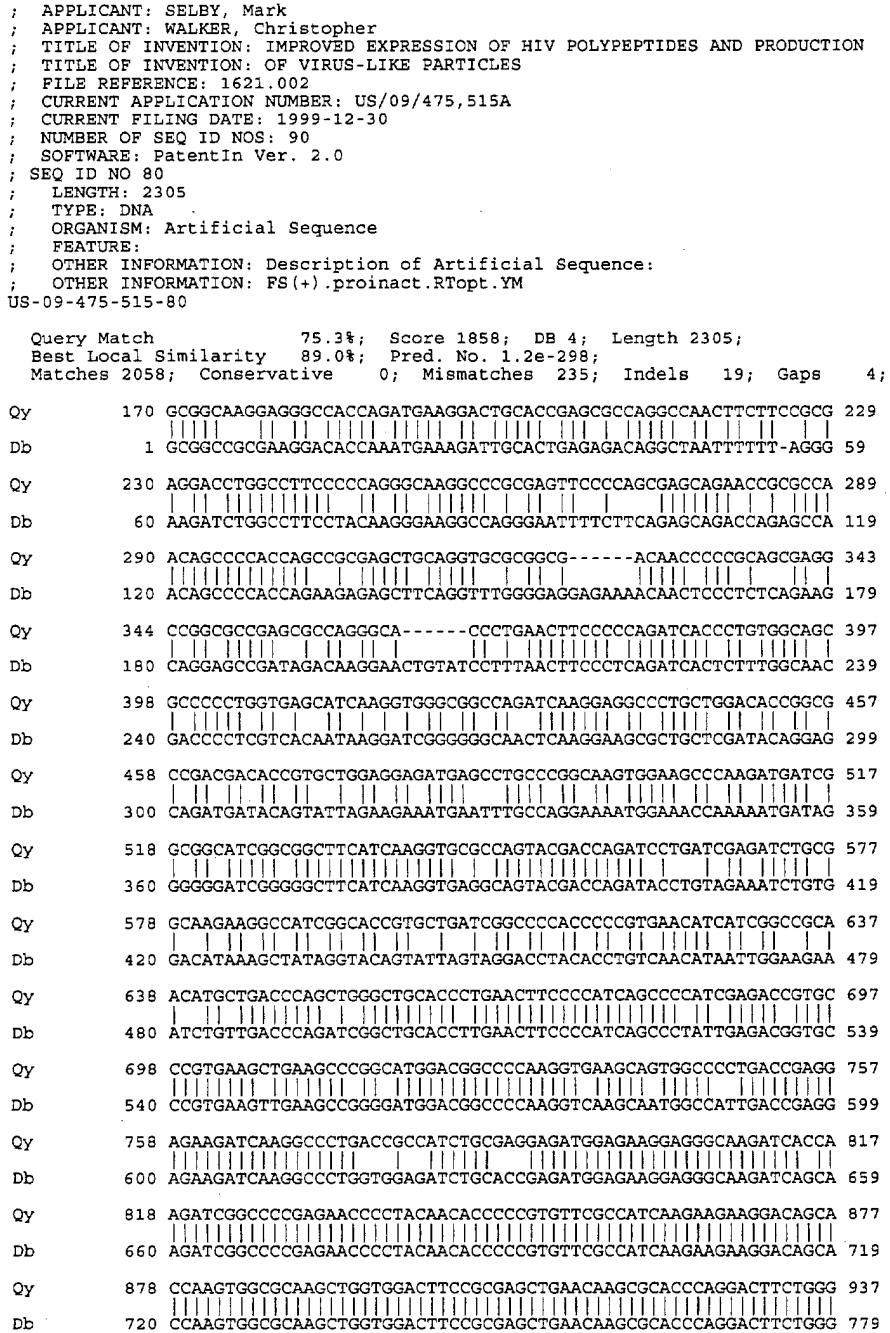


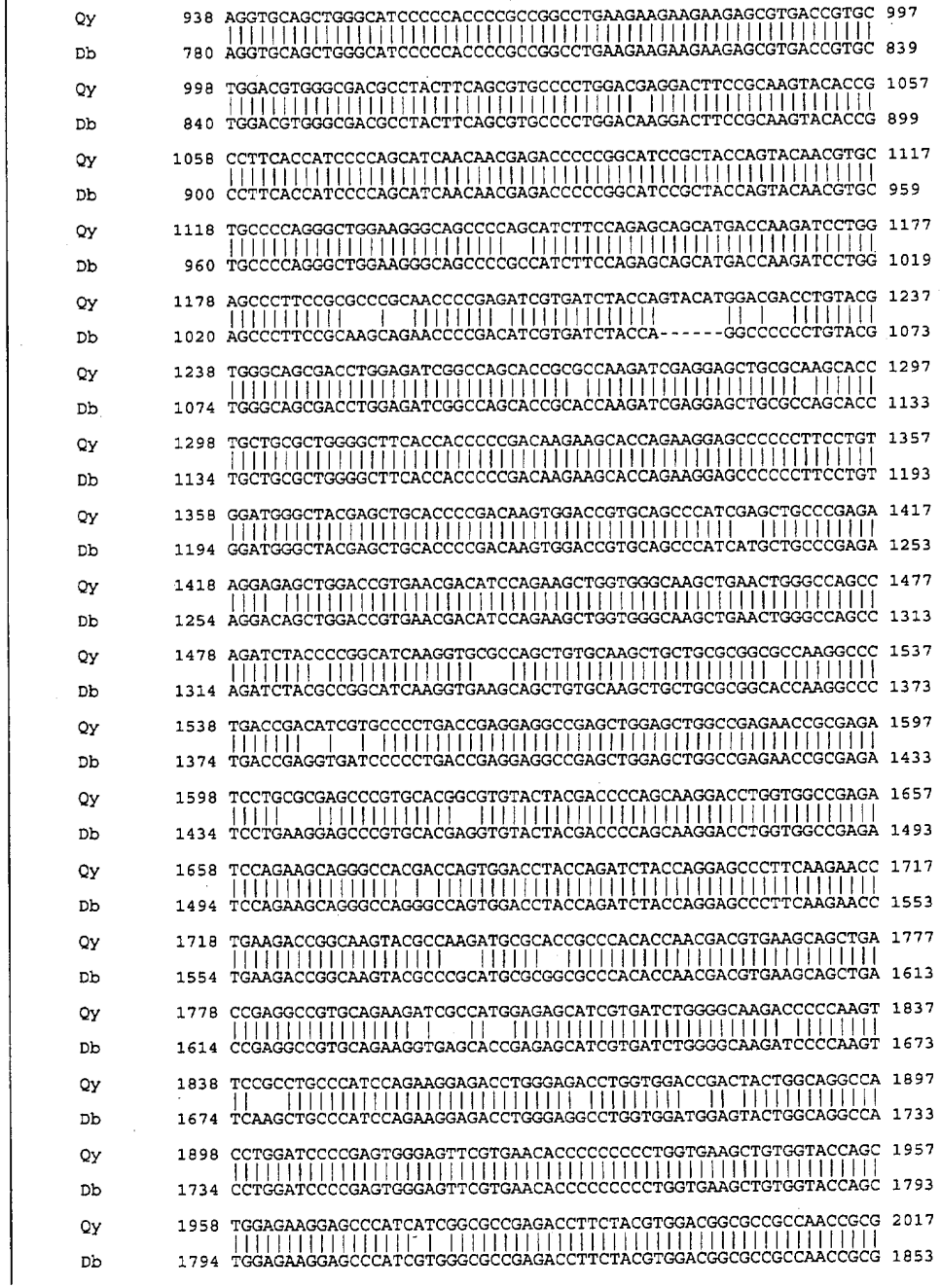


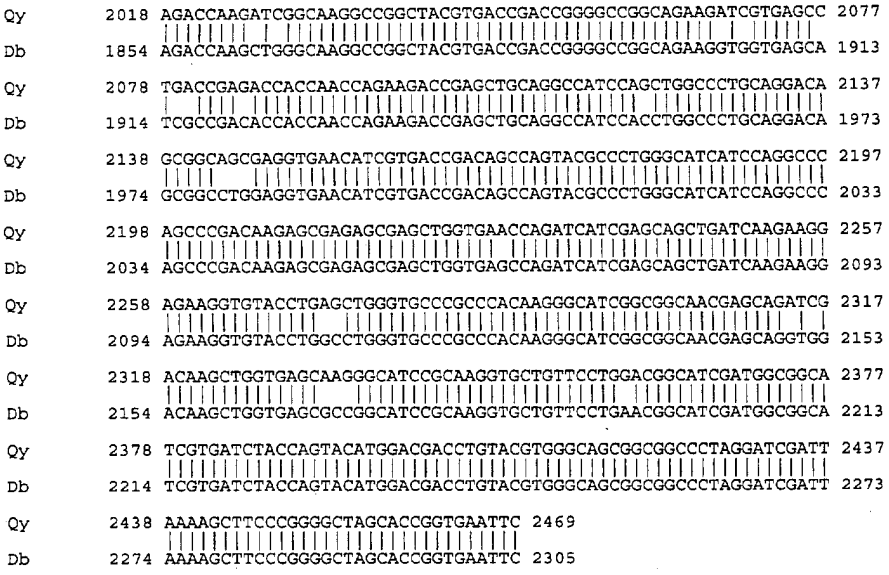
RESULT 5
US-09-475-515-80
; Sequence 80, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine

; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).proinact.RTopt.YM
US-09-475-515-80

Query Match 75.3%; Score 1858; DB 4; Length 2305;
Best Local Similarity 89.0%; Pred. No. 1.2e-298;
Matches 2058; Conservative 0; Mismatches 235; Indels 19; Gaps 4;

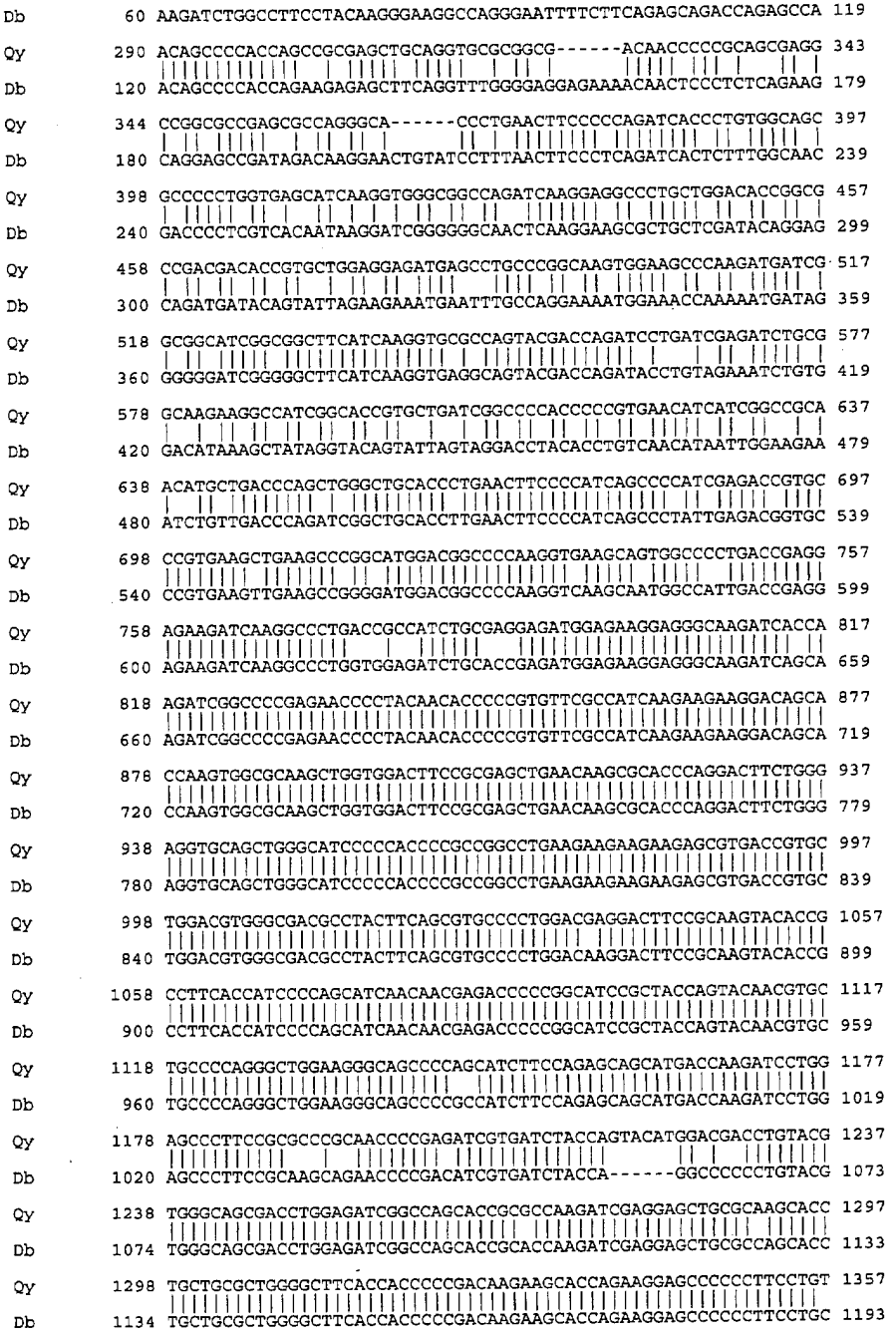






RESULT 6
US-09-475-515-81
; Sequence 81, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
; LENGTH: 2299
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).proinact.RTopt.YMWM
US-09-475-515-81

Query Match 74.2%; Score 1831.2; DB 4; Length 2299;
Best Local Similarity 88.6%; Pred. No. 3.1e-294;
Matches 2049; Conservative 0; Mismatches 238; Indels 25; Gaps 5;

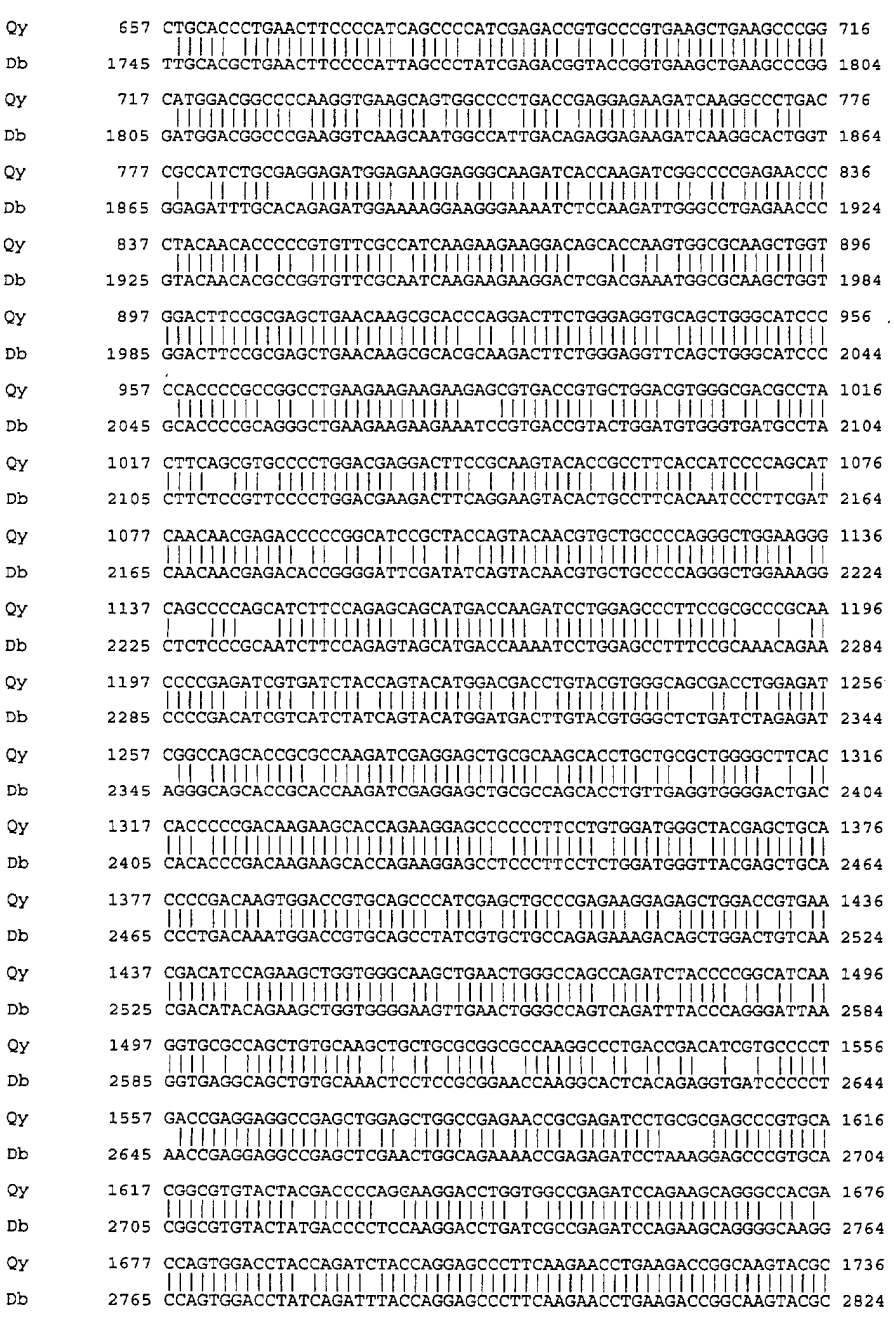


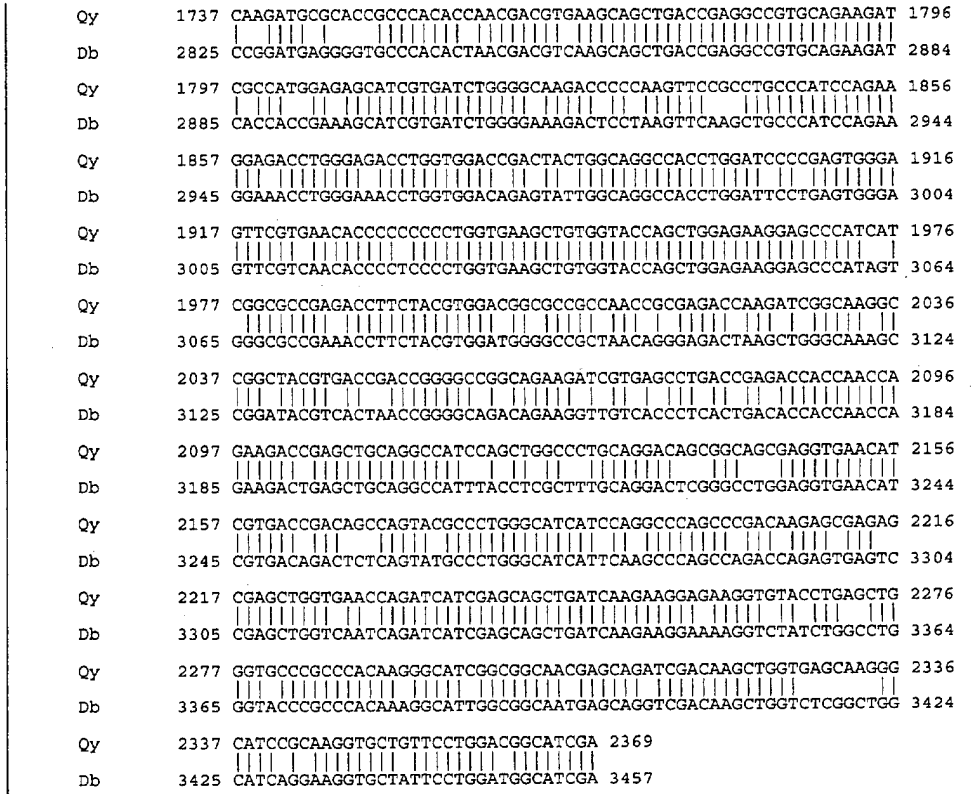
```
Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1194 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1247

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1248 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1307

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1308 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1367

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1368 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1427

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1428 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1487

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1488 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1547

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1548 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1607

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1608 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1667

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1668 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1727

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1728 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1787

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1788 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1847

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1848 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1907

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1908 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1967

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1968 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2027

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2028 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2087

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2088 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2147

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2148 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2207

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2208 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2267

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2268 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2299
```

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-SYNgp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 68.0%; Score 1678.6; DB 4; Length 4307;
Best Local Similarity 82.9%; Pred. No. 5.3e-269;
Matches 1968; Conservative 0; Mismatches 389; Indels 16; Gaps 4;

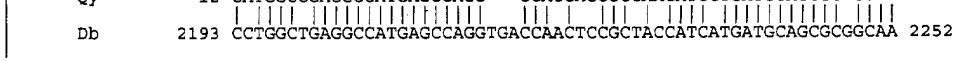
```
Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db       1086 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1145

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db       1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1205

Qy        129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db       1206 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCA 1265

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db       1266 GATGAAAGACTGTACGGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACA 1324

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db       1325 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCCGCGCCAACAGCCCCACCCGCCGCGA 1384

Qy        309 GCTGCAGGT------GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db       1385 GCTTCAGGTCTGGGGTCGCGACAACAACTCCCCCTCCGAAGCAGGAGCCGACCGCCAGGG 1444

Qy        363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db       1445 CACGGTGTCCTTCAACTTCCCTCAGGTCACGCTTTGGCAGCGACCCCTCGTCACCATCAA 1504

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db       1505 GATCGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1564

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db       1565 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1624

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db       1625 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1684

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db       1685 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1744
```

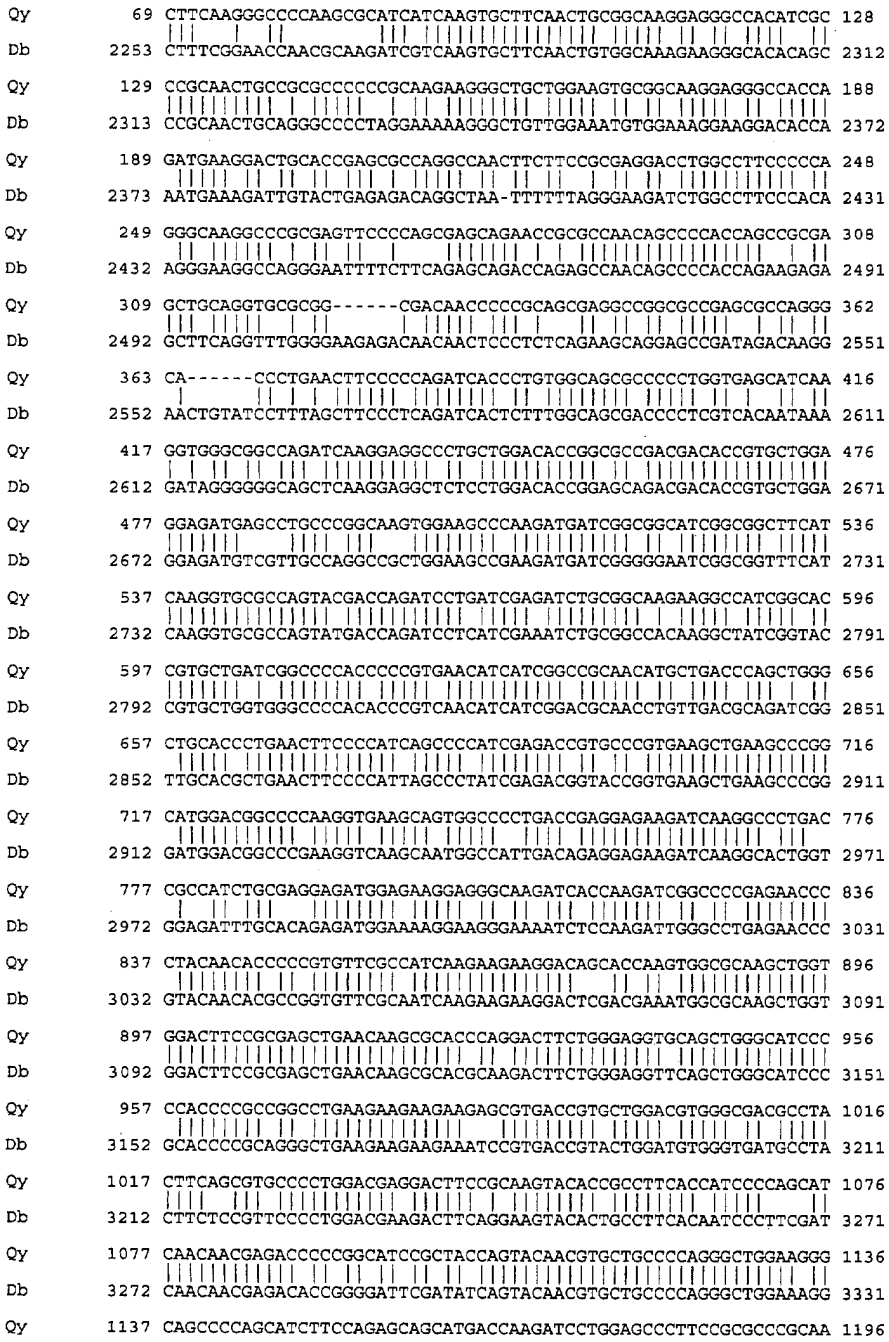


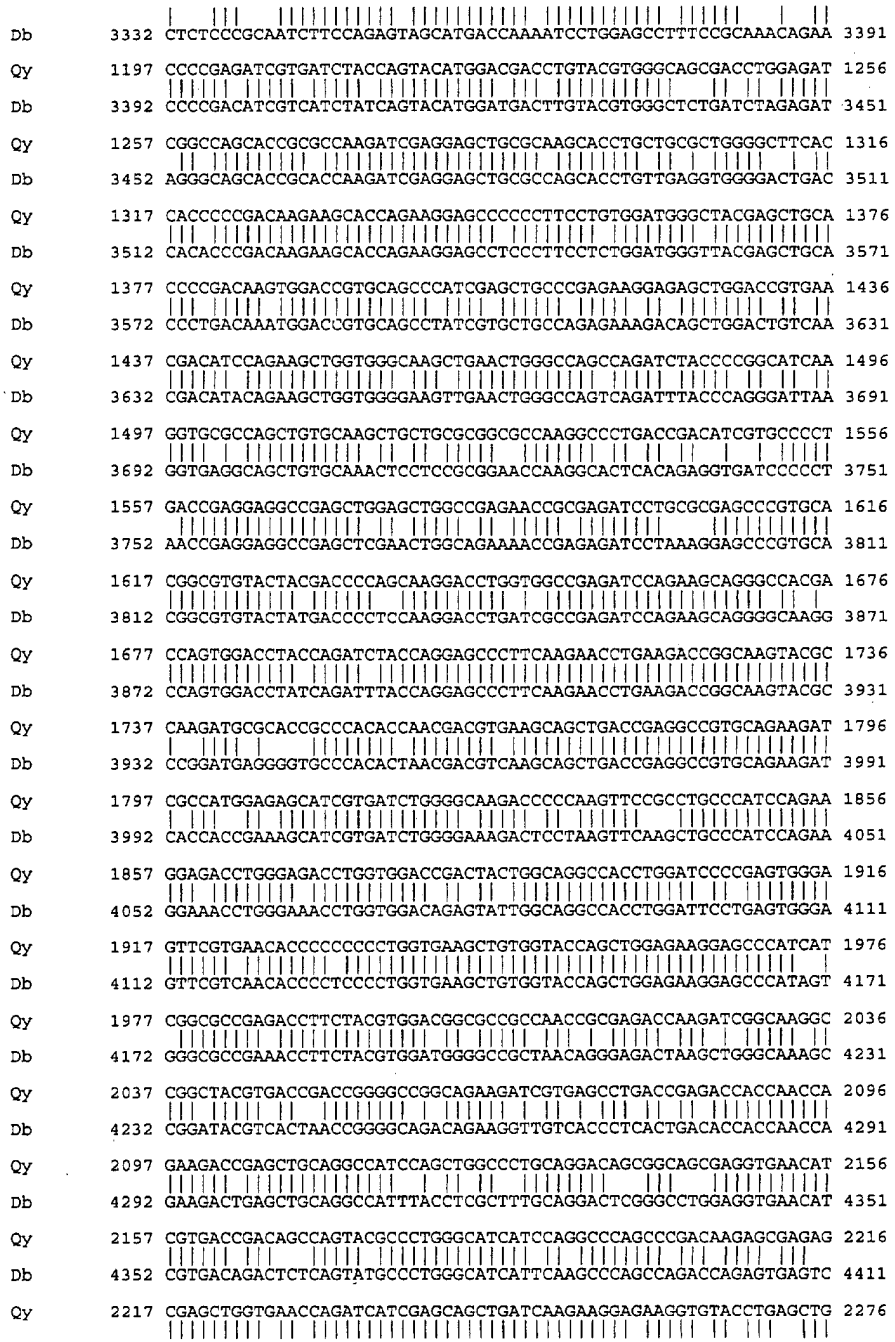


RESULT 8
US-09-552-950-5
; Sequence 5, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 9772
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:pSYNGP
US-09-552-950-5

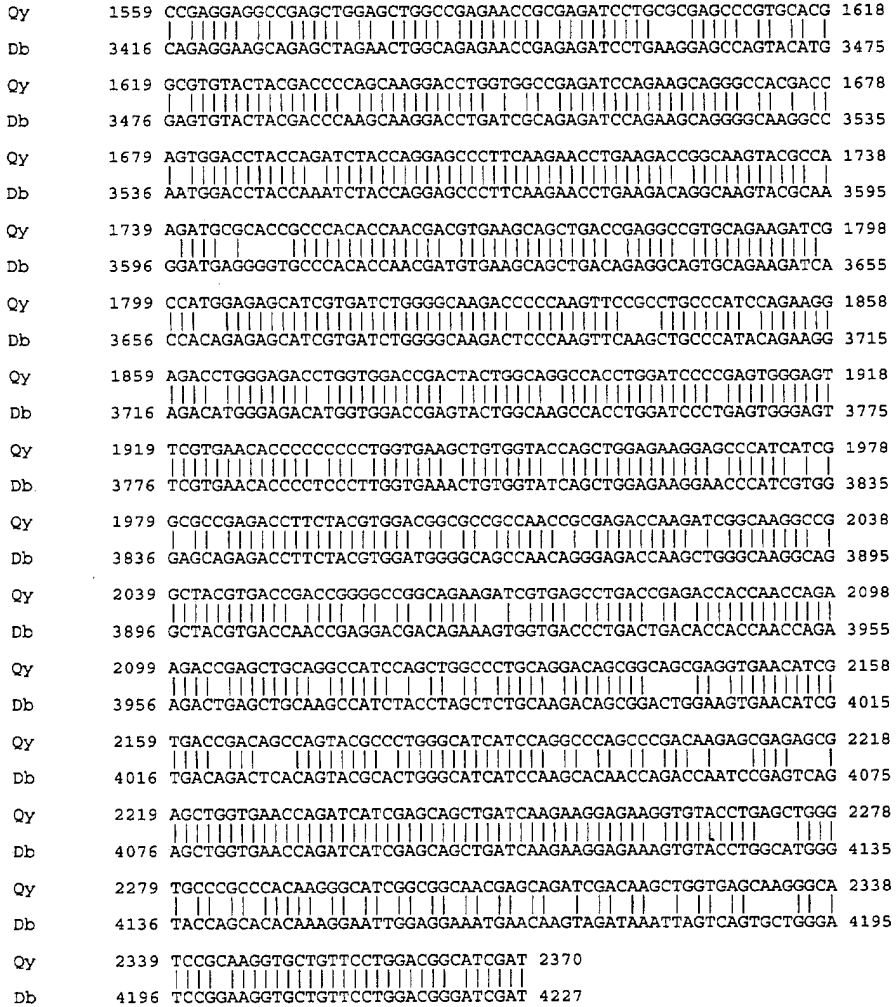
Query Match 66.9%; Score 1651.4; DB 4; Length 9772;
Best Local Similarity 82.2%; Pred. No. 1.8e-264;
Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;







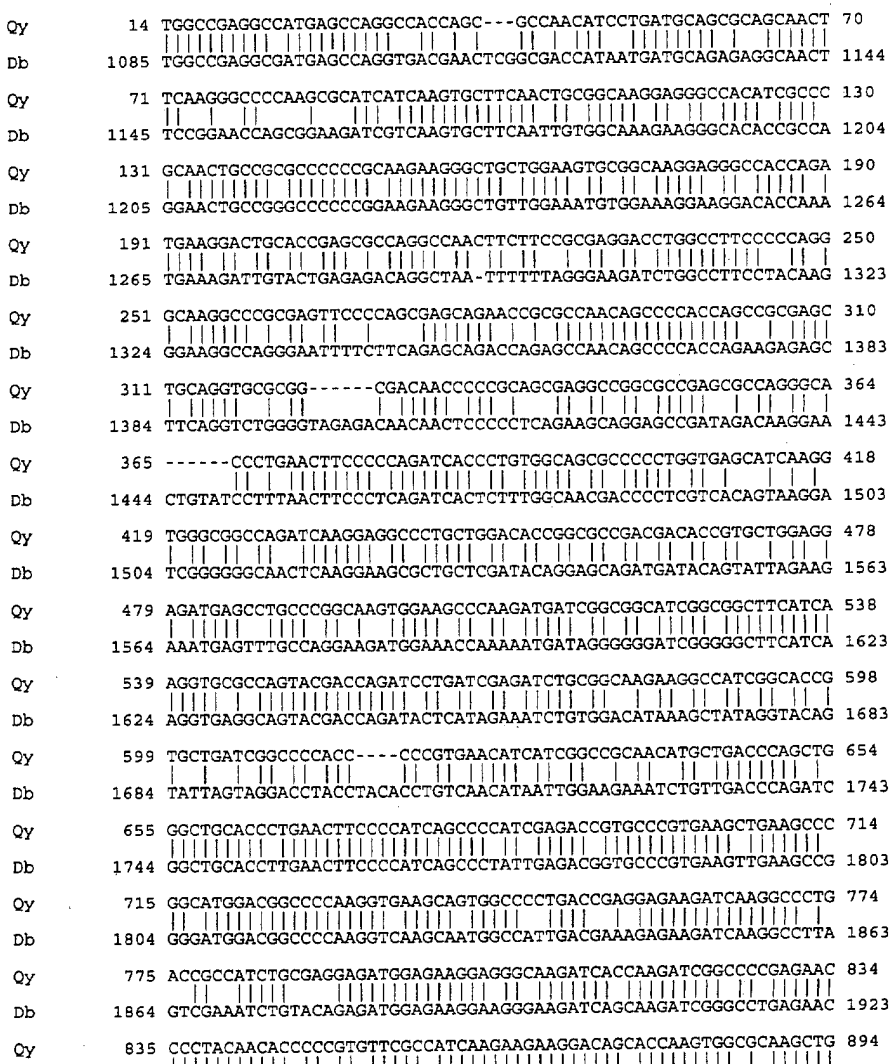
```
Db       4412 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 4471

Qy       2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336

Db       4472 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 4531

Qy       2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2369

Db       4532 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 4564


RESULT 9
US-09-872-733A-6
; Sequence 6, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733A
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;  LENGTH: 8366
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:  DNA sequence
;  OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
;  OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
;  OTHER INFORMATION: resistance gene
US-09-872-733A-6

  Query Match             64.5%;  Score 1592.8;  DB 4;  Length 8366;
  Best Local Similarity   80.7%;  Pred. No. 8.3e-255;
  Matches 1914;  Conservative    0;  Mismatches  442;  Indels   16;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70

Db       1857 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1916

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130

Db       1917 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1976

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190

Db       1977 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 2036

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250

Db       2037 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 2095

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310

Db       2096 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 2155

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364

Db       2156 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 2215

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418

Db       2216 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 2275

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478

Db       2276 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 2335

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538

Db       2336 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 2395

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598

Db       2396 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 2455

Qy        599 TGCTGATCGGCCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658

Db       2456 TATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCT 2515

Qy        659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718

Db       2516 GCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGA 2575

Qy        719 TGGACGGCCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778

Db       2576 TGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCG 2635

Qy        779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838

Db       2636 AAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCT 2695

Qy        839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898

Db       2696 ACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGG 2755

Qy        899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958

Db       2756 ACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCAC 2815

Qy        959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018

Db       2816 ATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACT 2875

Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078

Db       2876 TCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCA 2935

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138

Db       2936 ACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGAT 2995

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198

Db       2996 CACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACC 3055

Qy       1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258

Db       3056 CAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCG 3115

Qy       1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318

Db       3116 GGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCA 3175

Qy       1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378

Db       3176 CACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATC 3235

Qy       1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438

Db       3236 CTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACG 3295

Qy       1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498

Db       3296 ACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAG 3355

Qy       1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558

Db       3356 TTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCCACTGA 3415
```



RESULT 10
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

Query Match 63.1%; Score 1557.4; DB 4; Length 4338;
Best Local Similarity 80.5%; Pred. No. 5.4e-249;
Matches 1914; Conservative 0; Mismatches 441; Indels 22; Gaps 7;

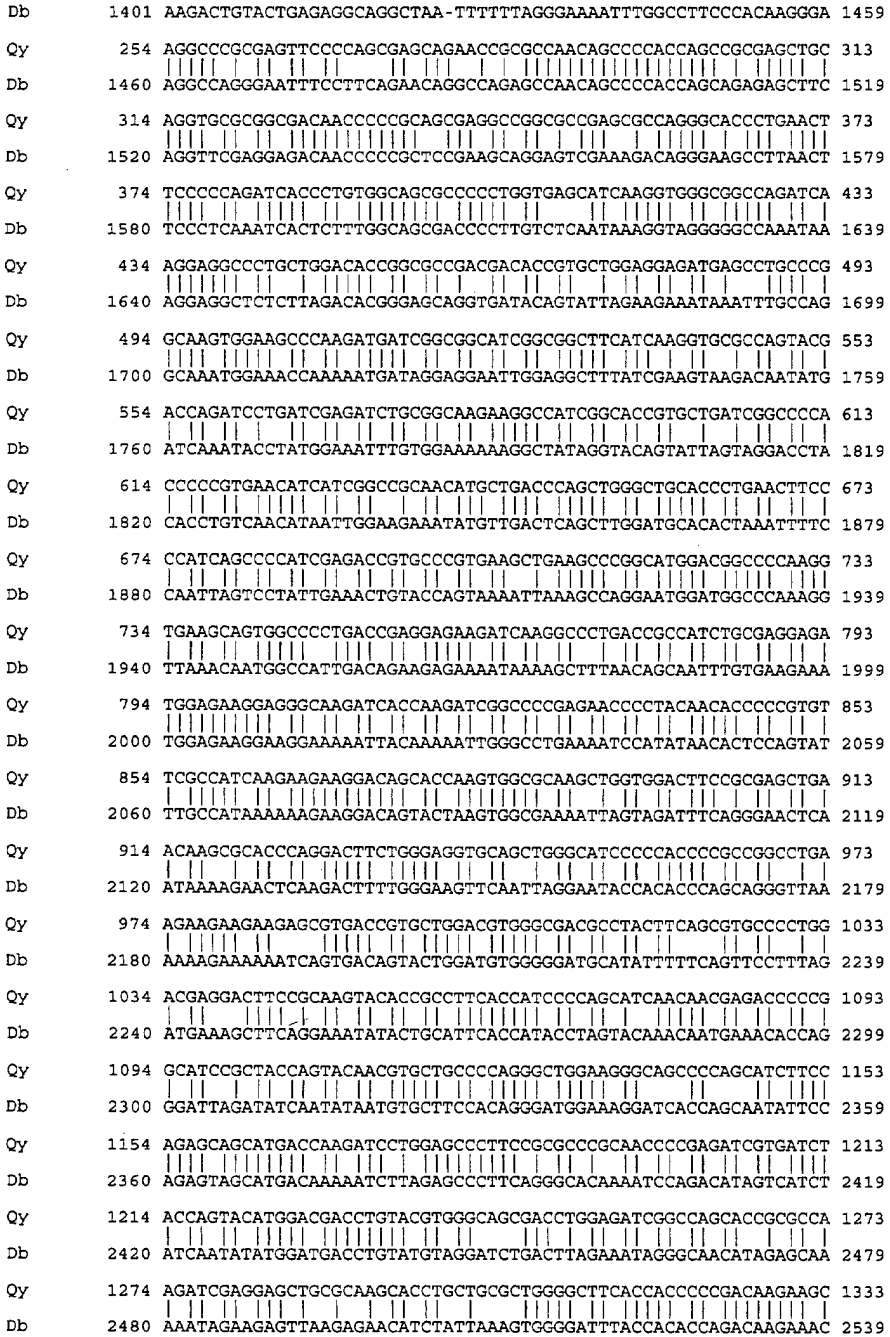


```
Db       1924 CCCTACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTG 1983
Qy        895 GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954
Db       1984 GTGGACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATC 2043
Qy        955 CCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC 1014
Db       2044 CCACATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCC 2103
Qy       1015 TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGC 1074
Db       2104 TACTTCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGC 2163
Qy       1075 ATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAG 1134
Db       2164 ATCAACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAG 2223
Qy       1135 GGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGC 1194
Db       2224 GGATCACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAA 2283
Qy       1195 AACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAG 1254
Db       2284 AACCCAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAG 2343
Qy       1255 ATC-GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTT 1313
Db       2344 ATCGGGGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACT 2403
Qy       1314 CACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCT 1373
Db       2404 GACCACACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACT 2463
Qy       1374 GCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGT 1433
Db       2464 GCATCCTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGT 2523
Qy       1434 GAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCAT 1493
Db       2524 GAACGACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCAT 2583
Qy       1494 CAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCC 1553
Db       2584 CAAAGTTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCC 2643
Qy       1554 CCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGT 1613
Db       2644 ACTGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGT 2703
Qy       1614 GCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCA 1673
Db       2704 ACATGGAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCA 2763
Qy       1674 CGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTA 1733
Db       2764 AGGCCAATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTA 2823
Qy       1734 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1793
Db       2824 CGCAAGGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAA 2883
Qy       1794 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1853
Db       2884 GATCACCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943
Qy       1854 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1913
Db       2944 GAAGGAGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTG 3003
Qy       1914 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1973
Db       3004 GGAGTTCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCAT 3063

Qy       1974 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2033
Db       3064 CGTGGGAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAA 3123
Qy       2034 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2093
Db       3124 GGCAGGCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAA 3183
Qy       2094 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2153
Db       3184 CCAGAAGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAA 3243
Qy       2154 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2213
Db       3244 CATCGTGACAGACTCACAGTACG-CATGGGCATCATCCAAGCACAACCAGACCAATCCGA 3302
Qy       2214 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2273
Db       3303 GTCAGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGC 3362
Qy       2274 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2333
Db       3363 ATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGC 3422
Qy       2334 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db       3423 TGGGATCCGGAAGGTGCTGTTCCTGGACGGGATCGAT 3459
```

RESULT 11
US-09-184-418C-8
; Sequence 8, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 8
; LENGTH: 9010
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=96ZM651; 137..1621:"gag"; 1426..4425:"pol";
; OTHER INFORMATION: 4370..4948:"vif"; 4888..5178:"vpr";
; OTHER INFORMATION: 5159..5373-7734..7824:"tat"; 5298..5373-7734..7981:"rev";
; OTHER INFORMATION: 5387..5647:"vpu"; 5565..8171:"env"; 8173..8793:"nef"
US-09-184-418C-8

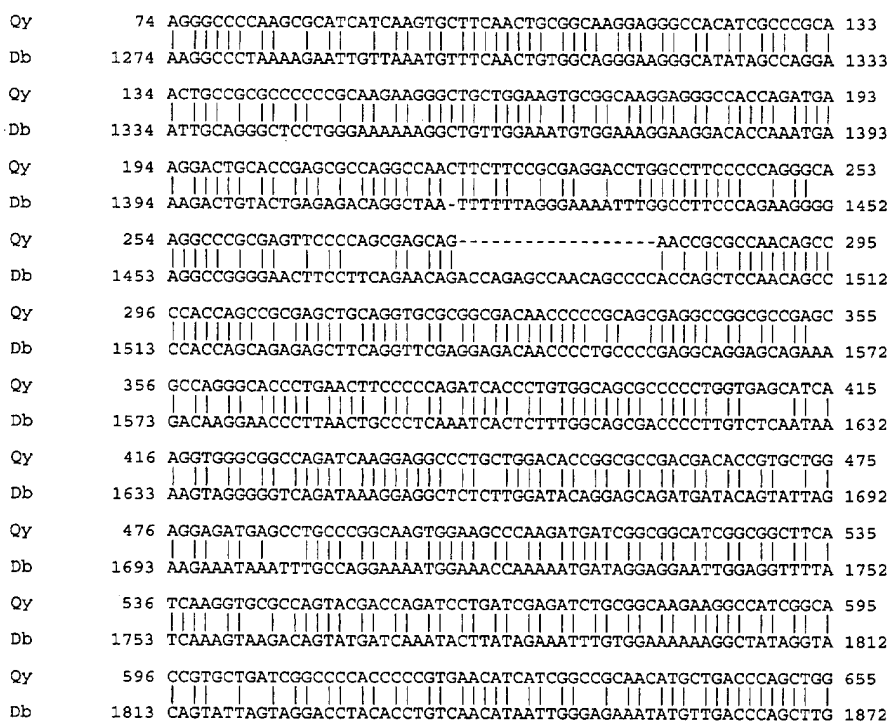
Query Match 49.7%; Score 1227.2; DB 4; Length 9010;
Best Local Similarity 70.3%; Pred. No. 2e-194;
Matches 1657; Conservative 0; Mismatches 698; Indels 1; Gaps 1;

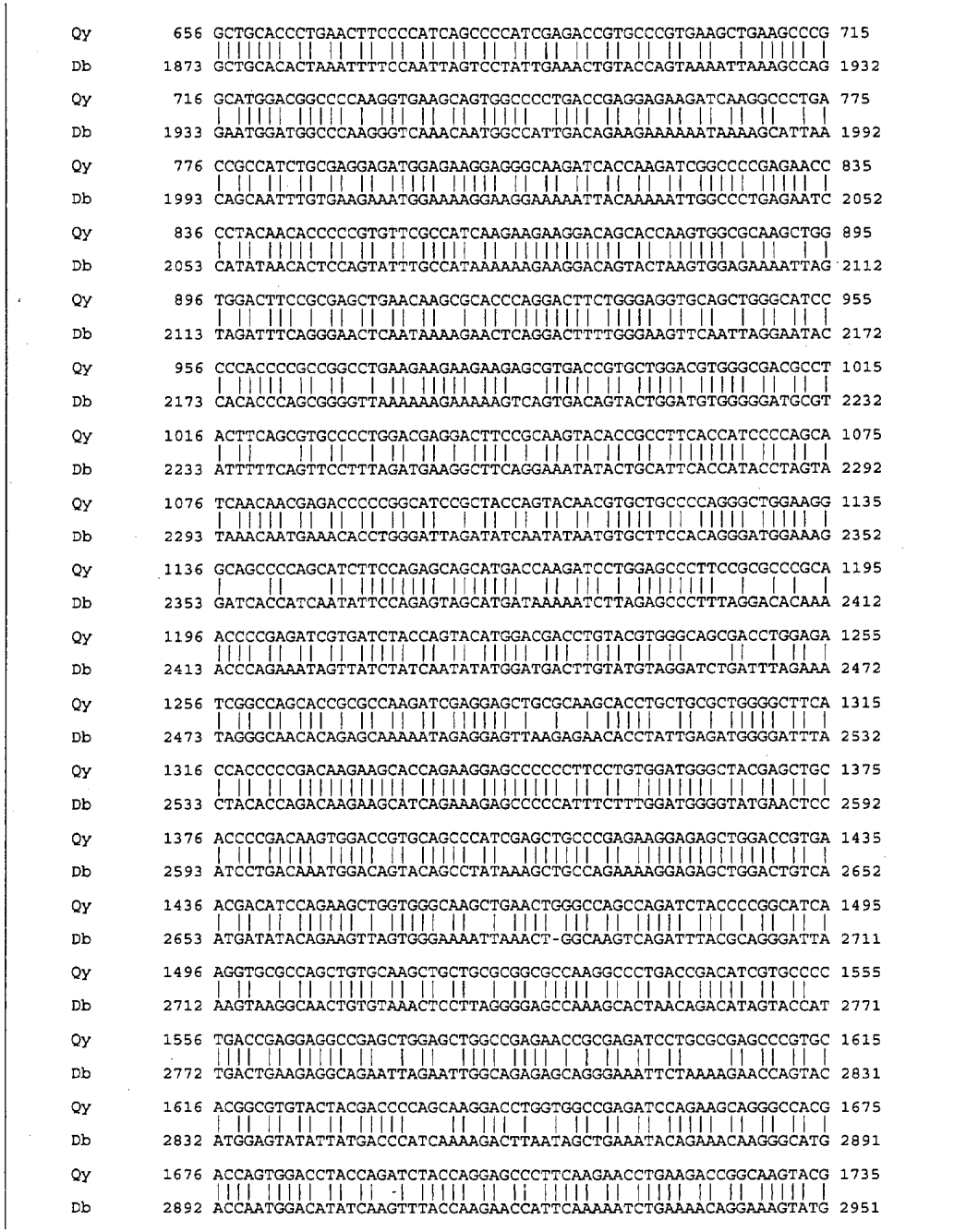
```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       1221 TGGCTGAGGCAATGAGCCAAACAAATAGTGTAAACATACTGATGCAGAAAAGCAATTTTA 1280
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       1281 AAGGAAATAAAAGAATGGTTAAATGTTTTAACTGTGGTAAGGAAGGGCACATAGCCAGAA 1340
Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       1341 ATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAGGGACACCAAATGA 1400
Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
```


Db 1401 AAGACTGTACTGAGAGGCAGGCTAA-TTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGA 1459
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 1460 AGGCCAGGGAATTTCCTTCAGAACAGGCCAGAGCCAACAGCCCCACCAGCAGAGAGCTTC 1519
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1520 AGGTTCGAGGAGACAACCCCCGCTCCGAAGCAGGAGTCGAAAGACAGGGAAGCCTTAACT 1579
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1580 TCCCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAGGTAGGGGGCCAAATAA 1639
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1640 AGGAGGCTCTCTTAGACACGGGAGCAGGTGATACAGTATTAGAAGAAATAAATTTGCCAG 1699
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1700 GCAAATGGAAACCAAAAATGATAGGAGGAATTGGAGGCTTTATCGAAGTAAGACAATATG 1759
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 1760 ATCAAATACCTATGGAAATTTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGACCTA 1819
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 1820 CACCTGTCAACATAATTGGAAGAAATATGTTGACTCAGCTTGGATGCACACTAAATTTTC 1879
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 1880 CAATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAGG 1939
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 1940 TTAAACAATGGCCATTGACAGAAGAGAAAATAAAAGCTTTAACAGCAATTTGTGAAGAAA 1999
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 2000 TGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGAAAATCCATATAACACTCCAGTAT 2059
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 2060 TTGCCATAAAAAAGAAGGACAGTACTAAGTGGCGAAAATTAGTAGATTTCAGGGAACTCA 2119
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 2120 ATAAAAGAACTCAAGACTTTTGGGAAGTTCAATTAGGAATACCACACCCAGCAGGGTTAA 2179
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2180 AAAAGAAAAAATCAGTGACAGTACTGGATGTGGGGGATGCATATTTTTCAGTTCCTTTAG 2239
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2240 ATGAAAGCTTCAGGAAATATACTGCATTCACCATACCTAGTACAAACAATGAAACACCAG 2299
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2300 GGATTAGATATCAATATAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCC 2359
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2360 AGAGTAGCATGACAAAAATCTTAGAGCCCTTCAGGGCACAAAATCCAGACATAGTCATCT 2419
Qy 1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db 2420 ATCAATATATGGATGACCTGTATGTAGGATCTGACTTAGAAATAGGGCAACATAGAGCAA 2479
Qy 1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db 2480 AAATAGAAGAGTTAAGAGAACATCTATTAAAGTGGGGATTTACCACACCAGACAAGAAAC 2539
Qy 1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db 2540 ATCAGAAAGAACCCCCATTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACAG 2599
Qy 1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db 2600 TACAGCCTATACAGCTGGCAGAAAAAGATAGCTGGACTGTTAATGATATACAGAAGTTAG 2659
Qy 1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db 2660 TGGGAAAATTAAACTGGGCAAGTCAGATTTACGCAGGGATTAAAGTAAGGCAACTTTGTA 2719
Qy 1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db 2720 AACTCCTTAGGGGAGCCAAAGCACTAACAGACATAGTACCACTAACTGAAGAAGCAGAAT 2779
Qy 1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db 2780 TAGAATTGGCAGAGAACAAGGAAATTTTAAAAGAACCAGTACATGGGGTATATTATGACC 2839
Qy 1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db 2840 CATCAAAAGACTTGATAGCTGAAATACAGAAACAAGGGCATGACCAATGGACATATCAAA 2899
Qy 1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db 2900 TTTACCAGGAACCATTCAAAAATCTGAAAACAGGGAAGTATGCAAAAATGAGGACAGCCC 2959
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db 2960 ACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCCCTGGAGAGCATAG 3019
Qy 1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db 3020 TAATATGGGGAAAGATTCCTAAATTTAGACTACCCATCCAAAAAGAAACATGGGAAACAT 3079
Qy 1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db 3080 GGTGGACAGACTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACCCCTC 3139
Qy 1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db 3140 TCTTAGTAAAATTATGGTACCAGCTGGAGAAAGAACCCATAGTAGGAGCAGAAACCTTCT 3199
Qy 1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db 3200 ATGTAGATGGAGCAGCCAATAGGGAAACTAAATTAGGAAAAGCAGGGTATATTACTGACA 3259
Qy 2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db 3260 GAGGAAGGCAAAAAATTGTTACTCTAACTGAAACAACAAATCAGAAGACTGAATTACAAG 3319
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db 3320 CAATTTACCTAGCTTTGCAAGATTCAGGATCAGAAGTAAACATAGTAACTGACTCACAGT 3379
Qy 2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db 3380 ATGCGTTAGGAATCATTCAAGCACATCCAGATAAGAGTGAATCAGAGTTAGTCAACCAAA 3439
Qy 2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
Db 3440 TAATAGAACAATTAATAAAGAAGGAAAGGGTCTACCTGTCATGGGTACCAGCACATAAAG 3499
Qy 2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db 3500 GAATTGGAGGTAATGAACAGGTAGATAAATTAGTAAGCAAGGGAATCAGGAAAGTGCTGT 3559
Qy 2354 TCCTGGACGGCATCGA 2369
Db 3560 TTCTAGATGGAATAGA 3575

RESULT 12
US-09-184-418C-9
; Sequence 9, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 9
; LENGTH: 8972
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=96ZM751.3; 137.1632:gag; 1419.4435:pol;
; OTHER INFORMATION: 4380.4958:vif; 4898.5188:vpr; 5169.7814:tat;
; OTHER INFORMATION: 5308.7938:rev; 5407.5667:vpu; 5585.8128:env;
; OTHER INFORMATION: 8130.8753:nef
US-09-184-418C-9

Query Match 48.4%; Score 1196.2; DB 4; Length 8972;
Best Local Similarity 70.0%; Pred. No. 2.6e-189;
Matches 1662; Conservative 0; Mismatches 693; Indels 20; Gaps 3;





```
Qy       1736 CCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGA 1795
Db       2952 CAAAAATGAGGACTGCCCACACTAATGATGTAAAACAGTTAACAGAGGCGGTGCAAAAAA 3011

Qy       1796 TCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGA 1855
Db       3012 TAGCCATGGAAAGCATAGTAATATGGGGAAAGATTCCTAAATTTAGGCTACCCATTCAAA 3071

Qy       1856 AGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGG 1915
Db       3072 AAGAAACATGGGAGACATGGTGGACAGACTATTGGCAAGCCACCTGGATTCCTGAGTGGG 3131

Qy       1916 AGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCA 1975
Db       3132 AGTTTGTTAATACTCCCCCCCTAGTAAAATTATGGTACCAGCTGGAGAAAGAACCCATAG 3191

Qy       1976 TCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGG 2035
Db       3192 CAGGAGCAGAAACTTACTATGTAGATGGAGCAGCCAATAGGGAAACTAAAATAGGAAAAG 3251

Qy       2036 CCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACC 2095
Db       3252 CAGGGTATGTTACTGACAGAGGAAGGCAAAAAATTGTTACTCTAACTGAAACAACAAATC 3311

Qy       2096 AGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACA 2155
Db       3312 AAAAGACTGAATTACAAGCAATTCAGTTAGCTTTGCAGGATTCAGGATCAGAAGTAAACA 3371

Qy       2156 TCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGA 2215
Db       3372 TAGTAACAGACTCACAGTATGCATTAGGAATCATCCAAGCACAACCAGATAAGAGTGAAT 3431

Qy       2216 GCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCT 2275
Db       3432 CAGAATTAGTCAATCAAATAATAGAACAGTTGATAAAAAAGGAAAGGGTTTACCTGTCAT 3491

Qy       2276 GGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGG 2335
Db       3492 GGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTGGTAAGTAGTG 3551

Qy       2336 GCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db       3552 GAATCAGGAAAGTGCTGTTTCTAGATGGAATAGAT 3586
```

RESULT 13
US-09-872-733A-3
; Sequence 3, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Pol gene
US-09-872-733A-3

Query Match 48.2%; Score 1190.6; DB 4; Length 2467;
Best Local Similarity 84.6%; Pred. No. 1.9e-188;

Matches 1337; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

```
Qy        790 GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCC 849
Db          7 GAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCTACAACACTCCA 66

Qy        850 GTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAG 909
Db         67 GTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAG 126

Qy        910 CTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGC 969
Db        127 CTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCCGCTGGG 186

Qy        970 CTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCC 1029
Db        187 TTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCCGTTCCC 246

Qy       1030 CTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACC 1089
Db        247 TTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAACGAGACA 306

Qy       1090 CCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATC 1149
Db        307 CCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATC 366

Qy       1150 TTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG 1209
Db        367 TTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACCCAGACATCGTG 426

Qy       1210 ATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGC 1269
Db        427 ATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGCACAGG 486

Qy       1270 GCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAG 1329
Db        487 ACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACACCAGACAAG 546

Qy       1330 AAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGG 1389
Db        547 AAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATCCTGACAAGTGG 606

Qy       1390 ACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAG 1449
Db        607 ACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACGACATACAGAAG 666

Qy       1450 CTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTG 1509
Db        667 CTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAGTTAGGCAGCTG 726

Qy       1510 TGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCC 1569
Db        727 TGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAGGAAGCA 786

Qy       1570 GAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTAC 1629
Db        787 GAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATGGAGTGTACTAC 846

Qy       1630 GACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTAC 1689
Db        847 GACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCCAATGGACCTAC 906

Qy       1690 CAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACC 1749
Db        907 CAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGGT 966

Qy       1750 GCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGC 1809
Db        967 GCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGC 1026

Qy       1810 ATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAG 1869
Db        1027 ATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACATGGGAG 1086
```

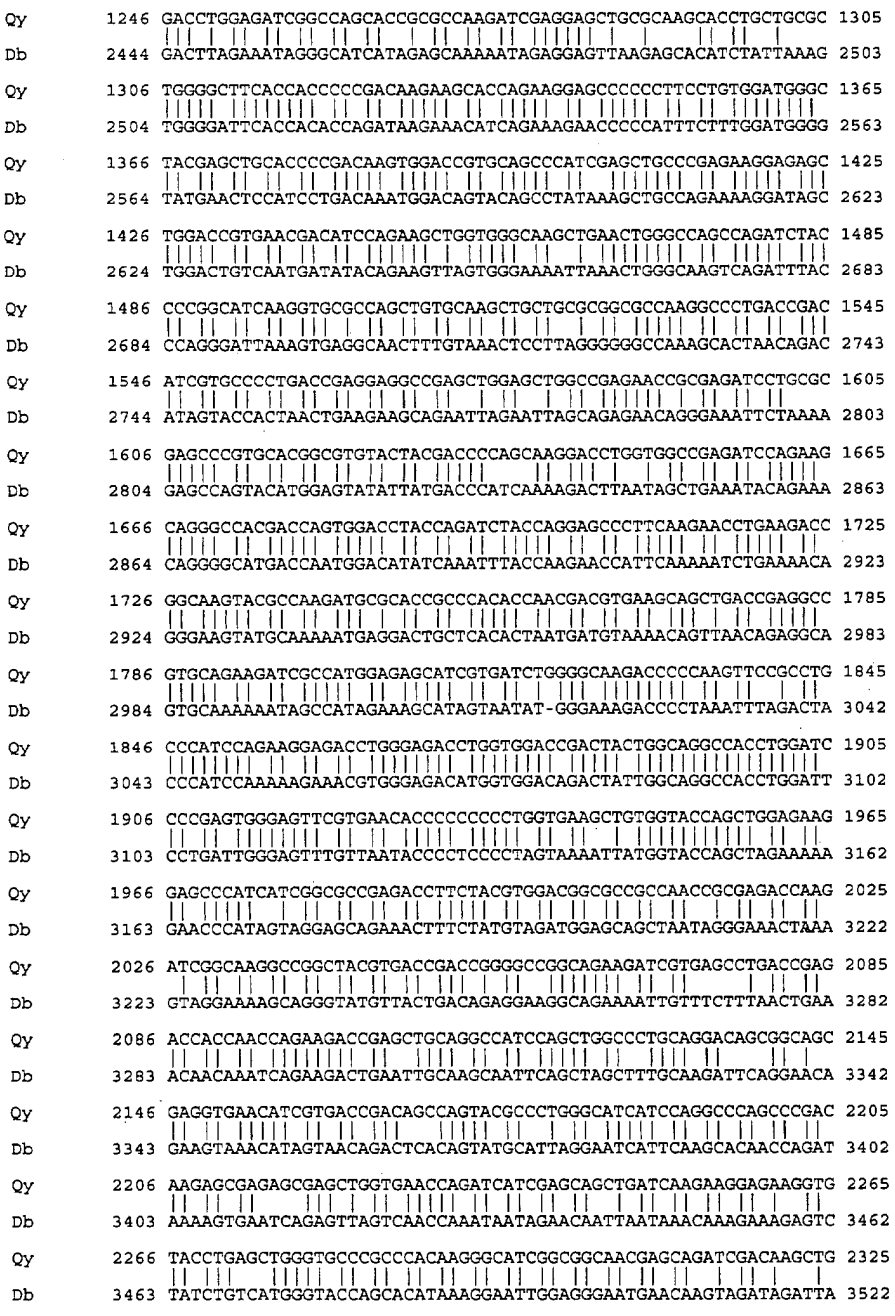
```
Qy       1870 ACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACC 1929
Db       1087 ACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTGAACACC 1146

Qy       1930 CCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACC 1989
Db       1147 CCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACC 1206

Qy       1990 TTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACC 2049
Db       1207 TTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTACGTGACC 1266

Qy       2050 GACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTG 2109
Db       1267 AACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTG 1326

Qy       2110 CAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGC 2169
Db       1327 CAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTGACAGACTCA 1386

Qy       2170 CAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAAC 2229
Db       1387 CAGTACGCACTGGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAGAGCTGGTGAAC 1446

Qy       2230 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCAC 2289
Db       1447 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGGTACCAGCACAC 1506

Qy       2290 AAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2349
Db       1507 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGATCCGGAAGGTG 1566

Qy       2350 CTGTTCCTGGACGGCATCGAT 2370
Db       1567 CTGTTCCTGGACGGGATCGAT 1587
```

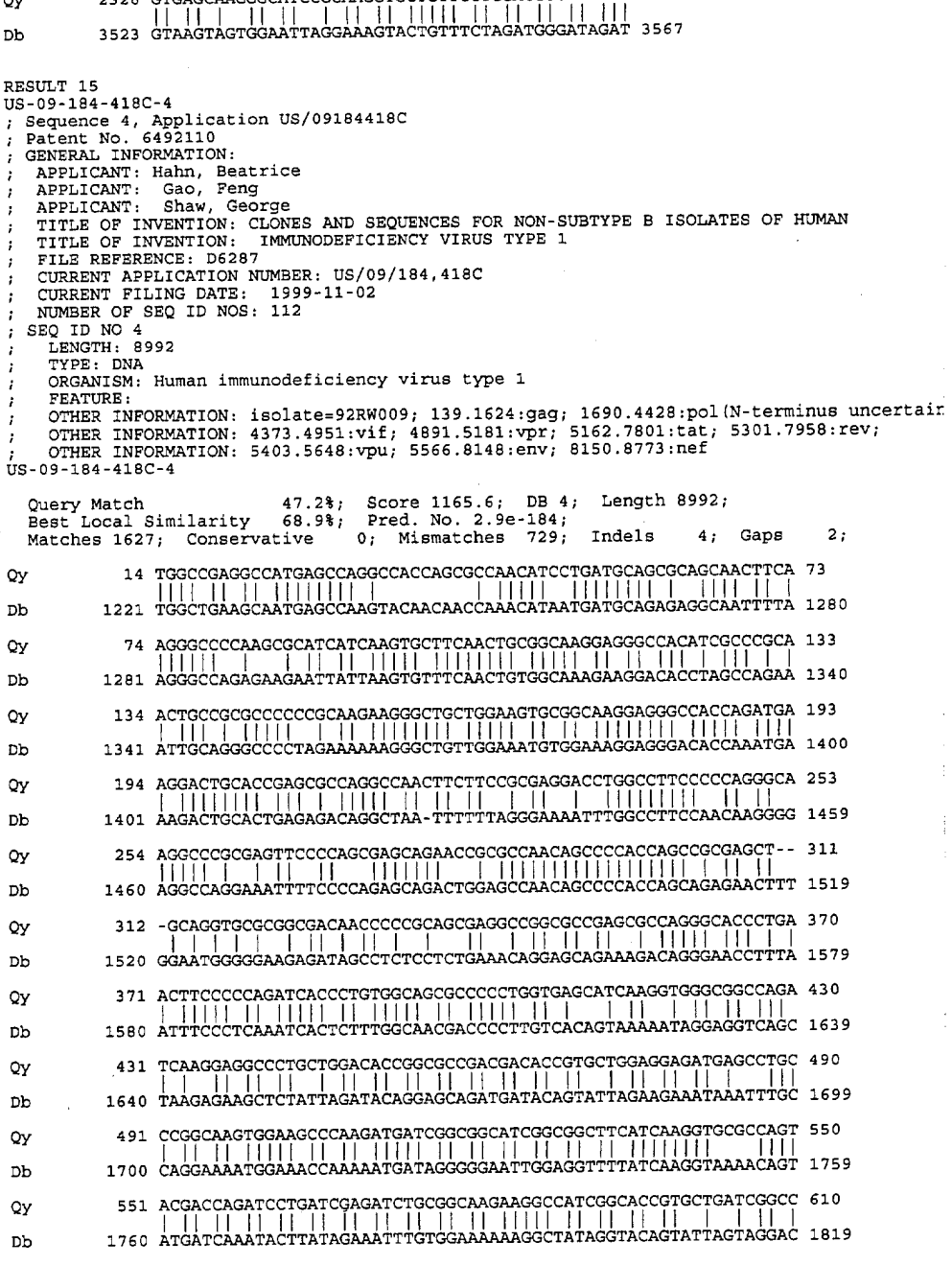
RESULT 14
US-09-184-418C-11
; Sequence 11, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 11
; LENGTH: 8959
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=94IN476.104; 138.1613:"gag";
; OTHER INFORMATION: 1418.4428:"pol"; 4361.4939:"vif"; 4879.5169:"vpr";
; OTHER INFORMATION: 5150.7782:tat"; 5289.7939:"rev"; 5378.5638:"vpu";
; OTHER INFORMATION: 5556.8129:"env"; 8131.8754:nef
US-09-184-418C-11

Query Match 48.2%; Score 1189; DB 4; Length 8959;
Best Local Similarity 70.0%; Pred. No. 4e-188;
Matches 1627; Conservative 0; Mismatches 696; Indels 2; Gaps 2;

```
Qy         46 AACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAAC 105
Db       1245 AACATAATGATGCAGAGAGGCAATTTTAAAGGCCCTAAAAGAATTGTTAAATGCTTCAAC 1304

Qy        106 TGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGG 165
```

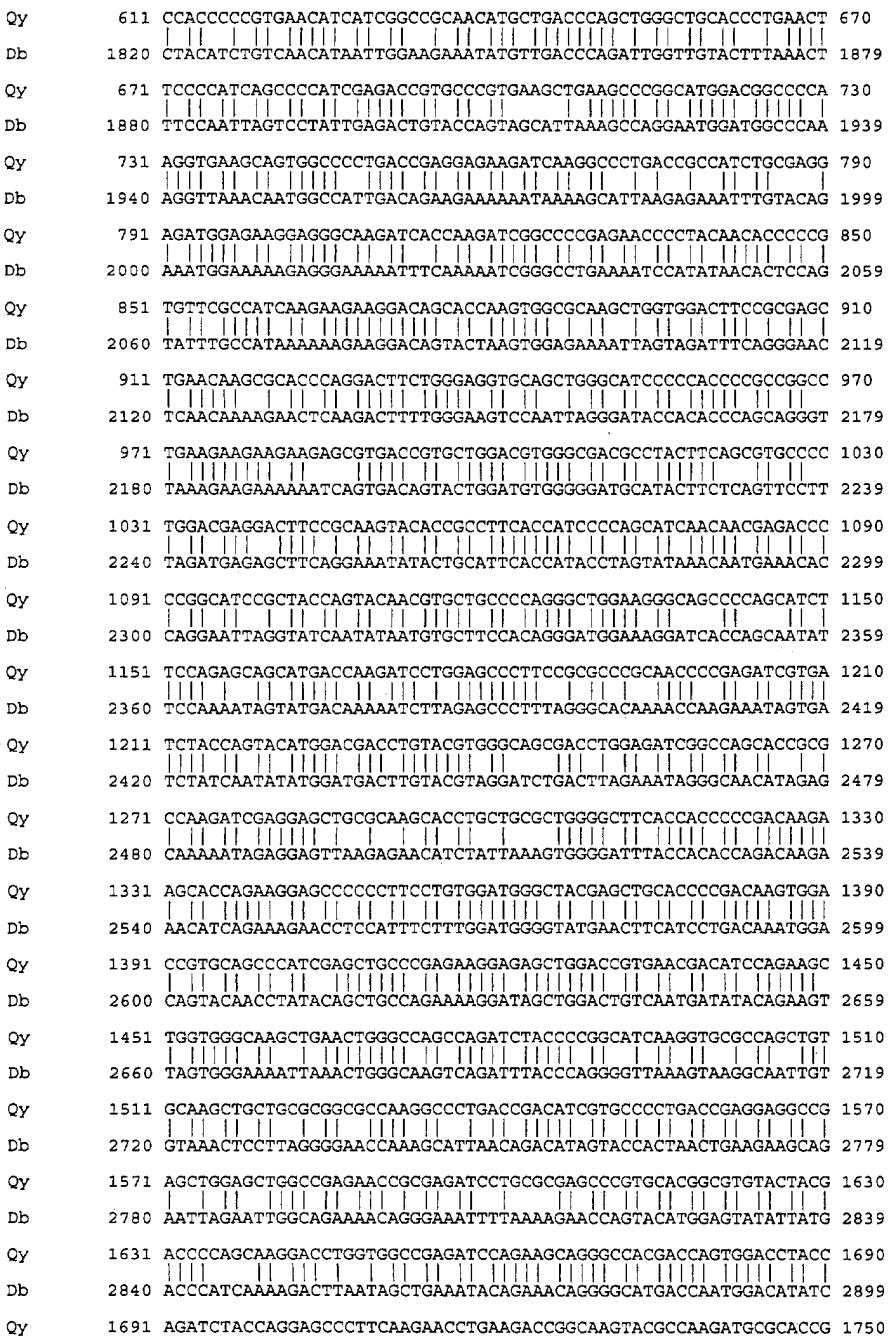
```
Db       1305 TGTGGCAAGGAAGGGCACATAGCCAGAAATTGCAGGGCCCCTAGAAAAAGAGGCTGTTGG 1364

Qy        166 AAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTC 225
Db       1365 AAATGTGGGCAAGAAGGACACCAAATGAAAGACTGTACTGAGAGGCAGGCTAA-TTTTTT 1423

Qy        226 CGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGC 285
Db       1424 AGGGAAAATTTGGCCTTCCCACAAGGGGAGGCCAGGGAATTTCCTTCAAAACAGGCCAGA 1483

Qy        286 GCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCC 345
Db       1484 GCCAACAGCCCCACCAGCAGAGAGCTTCAGGTTCAAGGAGACAACCCCCGCTCCGAAGCA 1543

Qy        346 GGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTG 405
Db       1544 GGAGTCGAAAGACAGGGAACCCTTAACTTCCCTCAAATCACTCTTTGGCAGCGACCCCTT 1603

Qy        406 GTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGAC 465
Db       1604 GTCTCAATAAAAGTAGGGGGCCAGATAAAGGAAGCTCTCTTAGACACAGGAGCAGATGAT 1663

Qy        466 ACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATC 525
Db       1664 ACAGTATTAGAAGAAATAGCTTTGCCAGGAAGATGGAAACCAAAAATGATAGGAGGAATT 1723

Qy        526 GGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAG 585
Db       1724 GGAGGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAG 1783

Qy        586 GCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTG 645
Db       1784 GCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAGATATGTTG 1843

Qy        646 ACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAG 705
Db       1844 ACTCAGCTTGGATGCACTCTAAATTTTCCAATTAGCCCCATTGAAACTGTACCAGTAAAA 1903

Qy        706 CTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATC 765
Db       1904 TTAAAGCCAGGAATGGATGGCCCAAAGGTTAAACAGTGGCCATTGACAGAAGAGAAAATA 1963

Qy        766 AAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGC 825
Db       1964 AAAGCATTAACAGAAATTTGTAAAGAAATGGAGAAGGAAGGAAAAATTACAAAAATTGGG 2023

Qy        826 CCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGG 885
Db       2024 CCTGAAAATCCATATAACACTCCAGTATTTGCCATAAAAAGGAAGGACAGTACTAAGTGG 2083

Qy        886 CGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAG 945
Db       2084 AGAAAATTAGTAGATTTCAGGGAGCTCAATAAAAGAACTCAAGACTTTTGGGAAGTTCAA 2143

Qy        946 CTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTG 1005
Db       2144 TTAGGAATACCACACCCAGCAGGTTTAAAAAAGAAAAAATCAGTGACAGTACTGGATGTG 2203

Qy       1006 GGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACC 1065
Db       2204 GGGGATGCATATTTTTCAGTTCCTTTAGATGAAGGCTTCGGGAAATATACTGCATTCACC 2263

Qy       1066 ATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAG 1125
Db       2264 ATACCTAGTATAAACAATGAAACACCAGGGATTAGATATCAATATAATGTGCTTCCACAG 2323

Qy       1126 GGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTC 1185
Db       2324 GGATGGAAAGGATCACCAGCAATATTCCAGAGTAGCATGACAAAAATCTTAGAGCCCTTT 2383

Qy       1186 CGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 1245
Db       2384 AGGGCACGAAATCCAAAAATAGTCATCTATCAATATATGGATGACTTGTATGTAGGGTCT 2443
```

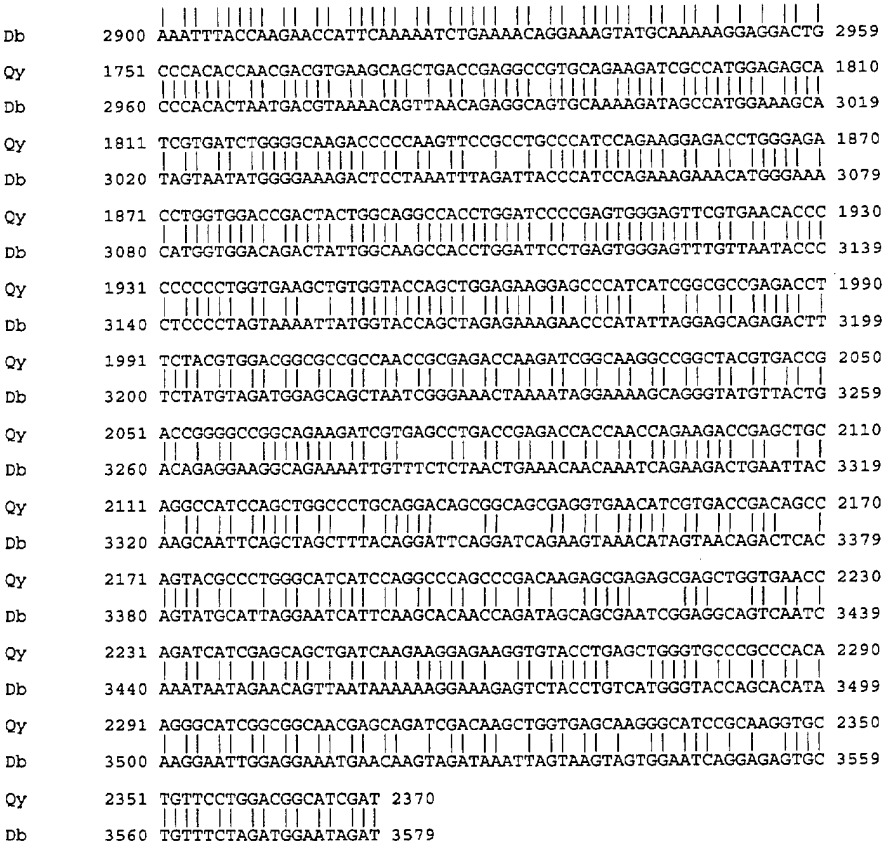




RESULT 15
US-09-184-418C-4
; Sequence 4, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 4
; LENGTH: 8992
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: isolate=92RW009; 139.1624:gag; 1690.4428:pol(N-terminus uncertain
; OTHER INFORMATION: 4373.4951:vif; 4891.5181:vpr; 5162.7801:tat; 5301.7958:rev;
; OTHER INFORMATION: 5403.5648:vpu; 5566.8148:env; 8150.8773:nef
US-09-184-418C-4

Query Match 47.2%; Score 1165.6; DB 4; Length 8992;
Best Local Similarity 68.9%; Pred. No. 2.9e-184;
Matches 1627; Conservative 0; Mismatches 729; Indels 4; Gaps 2;





Search completed: April 10, 2004, 16:22:37
Job time : 129.622 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 12:49:38 ; Search time 594.779 Seconds
(without alignments)
15572.780 Million cell updates/sec

Title: US-09-610-313-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2475585 seqs, 1875730760 residues

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

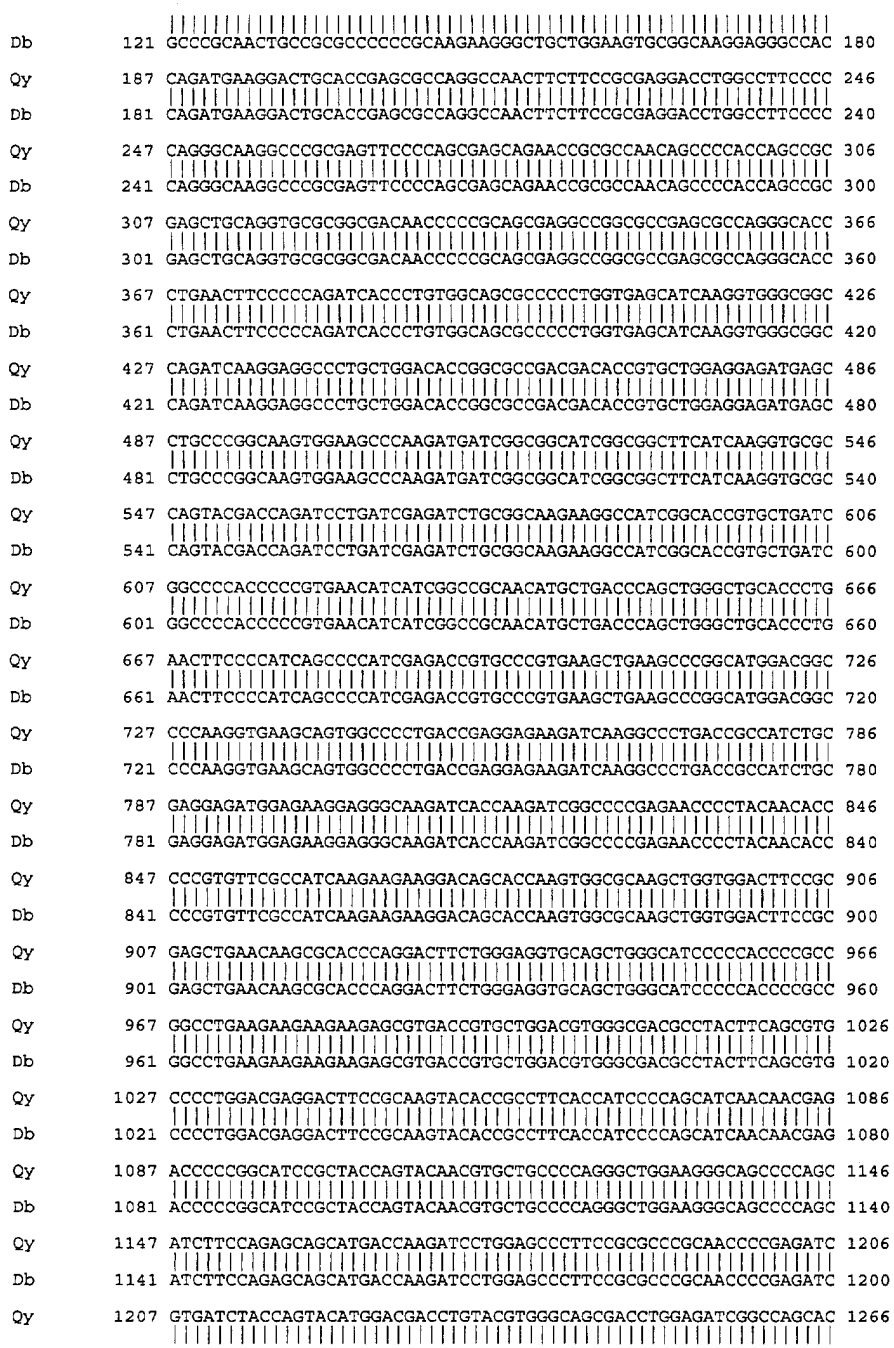
Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	99.5	2457	14	US-10-190-435-45	Sequence 45, Appl
2	2457	99.5	2457	14	US-10-190-305A-39	Sequence 39, Appl
3	2434.6	98.6	2457	14	US-10-190-435-44	Sequence 44, Appl
4	2434.6	98.6	2457	14	US-10-190-305A-38	Sequence 38, Appl
5	2401.8	97.3	2445	14	US-10-190-435-43	Sequence 43, Appl
6	2401.8	97.3	2445	14	US-10-190-305A-37	Sequence 37, Appl
7	2394.8	97.0	3930	14	US-10-190-435-9	Sequence 9, Appli
8	2393.2	96.9	3930	14	US-10-190-435-10	Sequence 10, Appl
9	2393.2	96.9	3930	14	US-10-190-435-11	Sequence 11, Appl
10	2393.2	96.9	5184	14	US-10-190-435-58	Sequence 58, Appl
11	2393.2	96.9	5184	14	US-10-190-305A-82	Sequence 82, Appl
12	2362.8	95.7	3531	14	US-10-190-435-13	Sequence 13, Appl
13	2361.2	95.6	3537	14	US-10-190-435-14	Sequence 14, Appl
14	2361.2	95.6	3537	14	US-10-190-435-15	Sequence 15, Appl
15	2360.2	95.6	5145	14	US-10-190-435-12	Sequence 12, Appl
16	2360.2	95.6	5145	14	US-10-190-305A-12	Sequence 12, Appl
17	2349.4	95.2	3624	14	US-10-190-435-47	Sequence 47, Appl
18	2349.4	95.2	3624	14	US-10-190-305A-41	Sequence 41, Appl
19	2301.8	93.2	3607	14	US-10-190-435-48	Sequence 48, Appl
20	2301.8	93.2	3607	14	US-10-190-305A-42	Sequence 42, Appl
21	2283.6	92.5	3597	14	US-10-190-435-46	Sequence 46, Appl
22	2283.6	92.5	3597	14	US-10-190-305A-40	Sequence 40, Appl
23	2149.6	87.1	2472	14	US-10-241-009-32	Sequence 32, Appl
24	2149.6	87.1	2472	14	US-10-190-434B-32	Sequence 32, Appl
25	2149.6	87.1	2472	14	US-10-190-305A-34	Sequence 34, Appl
26	2121.2	85.9	2466	14	US-10-241-009-31	Sequence 31, Appl
27	2121.2	85.9	2466	14	US-10-190-434B-31	Sequence 31, Appl
28	2121.2	85.9	2466	14	US-10-190-305A-33	Sequence 33, Appl
29	2094.4	84.8	2460	14	US-10-241-009-30	Sequence 30, Appl
30	2094.4	84.8	2460	14	US-10-190-434B-30	Sequence 30, Appl
31	2094.4	84.8	2460	14	US-10-190-305A-32	Sequence 32, Appl
32	2093.4	84.8	3564	14	US-10-241-009-13	Sequence 13, Appl
33	2093.4	84.8	3564	14	US-10-241-009-14	Sequence 14, Appl
34	2093.4	84.8	3564	14	US-10-190-434B-13	Sequence 13, Appl
35	2093.4	84.8	3564	14	US-10-190-434B-14	Sequence 14, Appl
36	2092.8	84.8	4716	14	US-10-190-435-17	Sequence 17, Appl
37	2092.8	84.8	4716	14	US-10-190-305A-13	Sequence 13, Appl
38	2089	84.6	3999	14	US-10-241-009-9	Sequence 9, Appli
39	2089	84.6	3999	14	US-10-190-434B-9	Sequence 9, Appli
40	2087.4	84.5	3999	14	US-10-241-009-10	Sequence 10, Appl
41	2087.4	84.5	3999	14	US-10-241-009-11	Sequence 11, Appl
42	2087.4	84.5	3999	14	US-10-190-434B-10	Sequence 10, Appl
43	2087.4	84.5	3999	14	US-10-190-434B-11	Sequence 11, Appl
44	2087.4	84.5	5283	14	US-10-241-009-54	Sequence 54, Appl
45	2087.4	84.5	5283	14	US-10-190-434B-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

Query Match 99.5%; Score 2457; DB 14; Length 2457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
```


Db 121 GCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180
Qy 187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db 181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240
Qy 247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db 241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300
Qy 307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db 301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360
Qy 367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db 361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420
Qy 427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db 421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480
Qy 487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db 481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540
Qy 547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db 541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600
Qy 607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db 601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660
Qy 667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db 661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720
Qy 727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780
Qy 787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db 781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840
Qy 847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db 841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy 907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db 901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy 967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db 961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy 1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db 1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy 1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db 1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy 1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
Db 1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy 1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy 1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
Db 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy 1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy 1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
Db 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy 1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
Db 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy 1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
Db 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy 1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
Db 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy 1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
Db 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy 1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
Db 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy 1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy 1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
Db 1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy 1927 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
Db 1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy 1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
Db 1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy 2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
Db 2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy 2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
Db 2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy 2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
Db 2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy 2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
Db 2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy 2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
Db 2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

```
Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 2
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

  Query Match             99.5%;  Score 2457;  DB 14;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
```

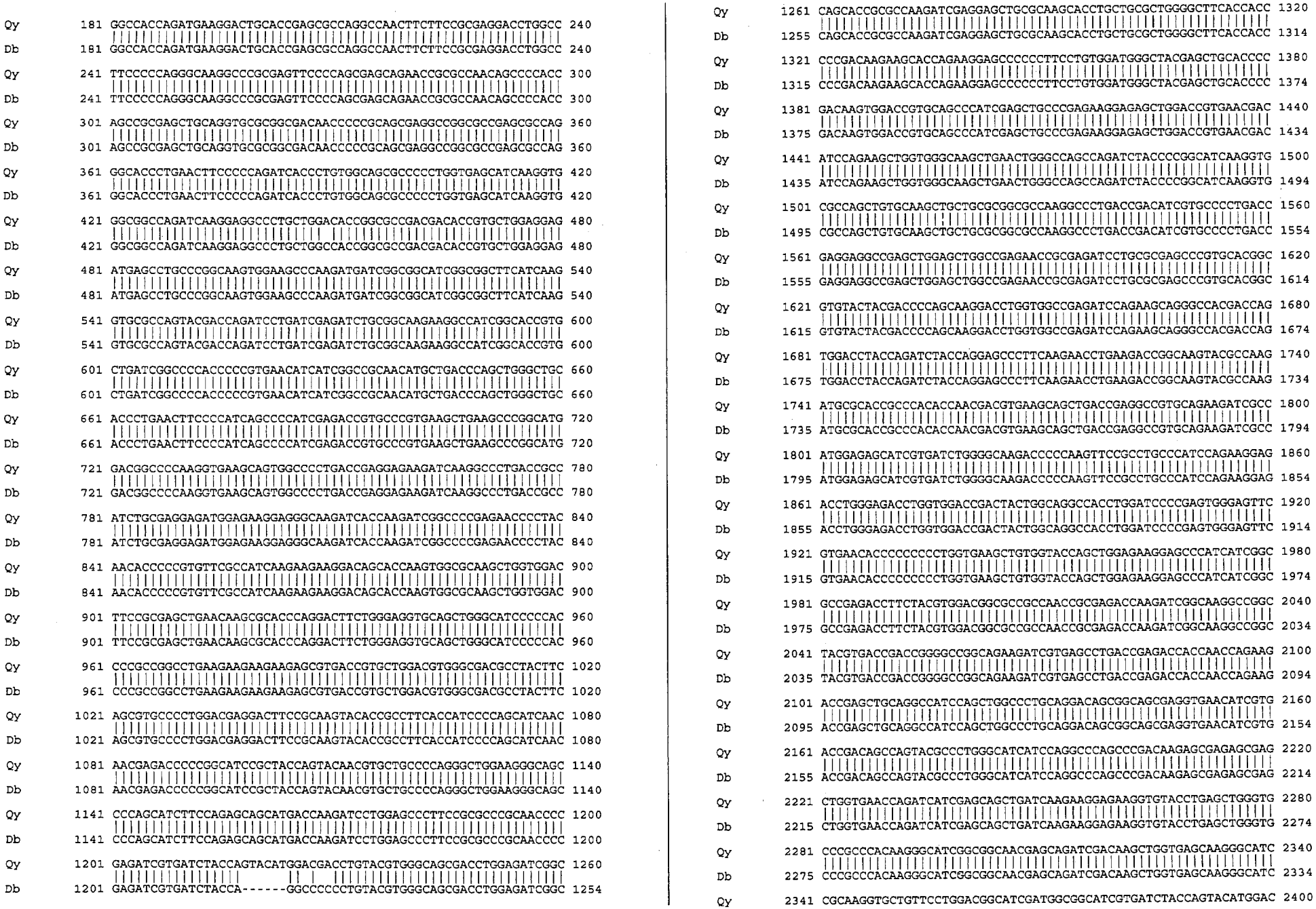
```
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746

Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806

Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866

Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926

Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1927 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986

Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046

Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106

Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166

Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226

Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286

Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346

Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406

Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 3
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 44
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44

  Query Match             98.6%;  Score 2434.6;  DB 14;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2453;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
```

```
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||  || |  ||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463
              |||
Db       2455 GGT 2457
```

RESULT 4
US-10-190-305A-38
; Sequence 38, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-305A-38

Query Match 98.6%; Score 2434.6; DB 14; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
```



```
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463

Db       2455 GGT 2457
```

RESULT 5
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMWM_C
US-10-190-435-43

Query Match 97.3%; Score 2401.8; DB 14; Length 2445;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66

Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126

Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186

Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246

Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306

Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366

Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426

Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486

Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546

Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606

Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666

Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726

Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786

Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846

Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906

Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966

Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026

Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086

Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146

Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206

Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266

Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326

Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386

Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446

Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506

Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566

Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
```

```
Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686

Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746

Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806

Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866

Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926

Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908

Qy       1927 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986

Db       1909 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046

Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106

Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166

Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226

Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCC 2286

Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346

Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406

Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 6
US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: p2Pol.opt.YMWM_C
US-10-190-305A-37

Query Match 97.3%; Score 2401.8; DB 14; Length 2445;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

```
Qy         7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66

Db         1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy        67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126

Db        61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy       127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186

Db       121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy       187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246

Db       181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy       247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306

Db       241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy       307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366

Db       301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy       367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426

Db       361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy       427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486

Db       421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy       487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546

Db       481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy       547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606

Db       541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy       607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666

Db       601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy       667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726

Db       661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy       727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786

Db       721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy       787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846

Db       781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy       847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
```

```
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966

Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026

Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086

Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146

Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206

Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266

Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326

Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386

Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446

Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506

Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566

Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626

Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686

Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746

Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806

Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866

Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926

Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986

Db       1909 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046

Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106

Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166

Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226

Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286

Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346

Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406

Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 7
US-10-190-435-9
; Sequence 9, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut_C
US-10-190-435-9

  Query Match   97.0%; Score 2394.8; DB 14; Length 3930;
  Best Local Similarity 99.2%; Pred. No. 0;
  Matches 2431; Conservative   0; Mismatches   7; Indels   12; Gaps   2;

```
Qy          14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db        1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy          74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db        1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy         134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
```

```
Db      1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666
Qy       194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db      1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726
Qy       254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db      1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786
Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db      1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db      1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906
Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db      1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966
Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db      1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026
Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db      2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db      2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db      2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db      2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy      1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db      2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740
```

```
Qy      1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db      2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800
Qy      1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db      2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854
Qy      1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db      2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914
Qy      1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db      2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974
Qy      1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db      2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034
Qy      1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db      3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094
Qy      1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db      3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154
Qy      1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db      3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214
Qy      1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db      3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274
Qy      1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db      3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334
Qy      1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db      3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394
Qy      1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db      3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454
Qy      1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db      3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514
Qy      2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db      3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574
Qy      2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db      3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634
Qy      2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db      3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694
Qy      2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
Db      3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754
Qy      2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db      3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814
```

```
Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 8
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

  Query Match             96.9%;  Score 2393.2;  DB 14;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||     || | ||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              ||||||||||||||||||||  ||  ||||||||||||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094
```

```
Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 9
US-10-190-435-11
; Sequence 11, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna_C
US-10-190-435-11

Query Match 96.9%; Score 2393.2; DB 14; Length 3930;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

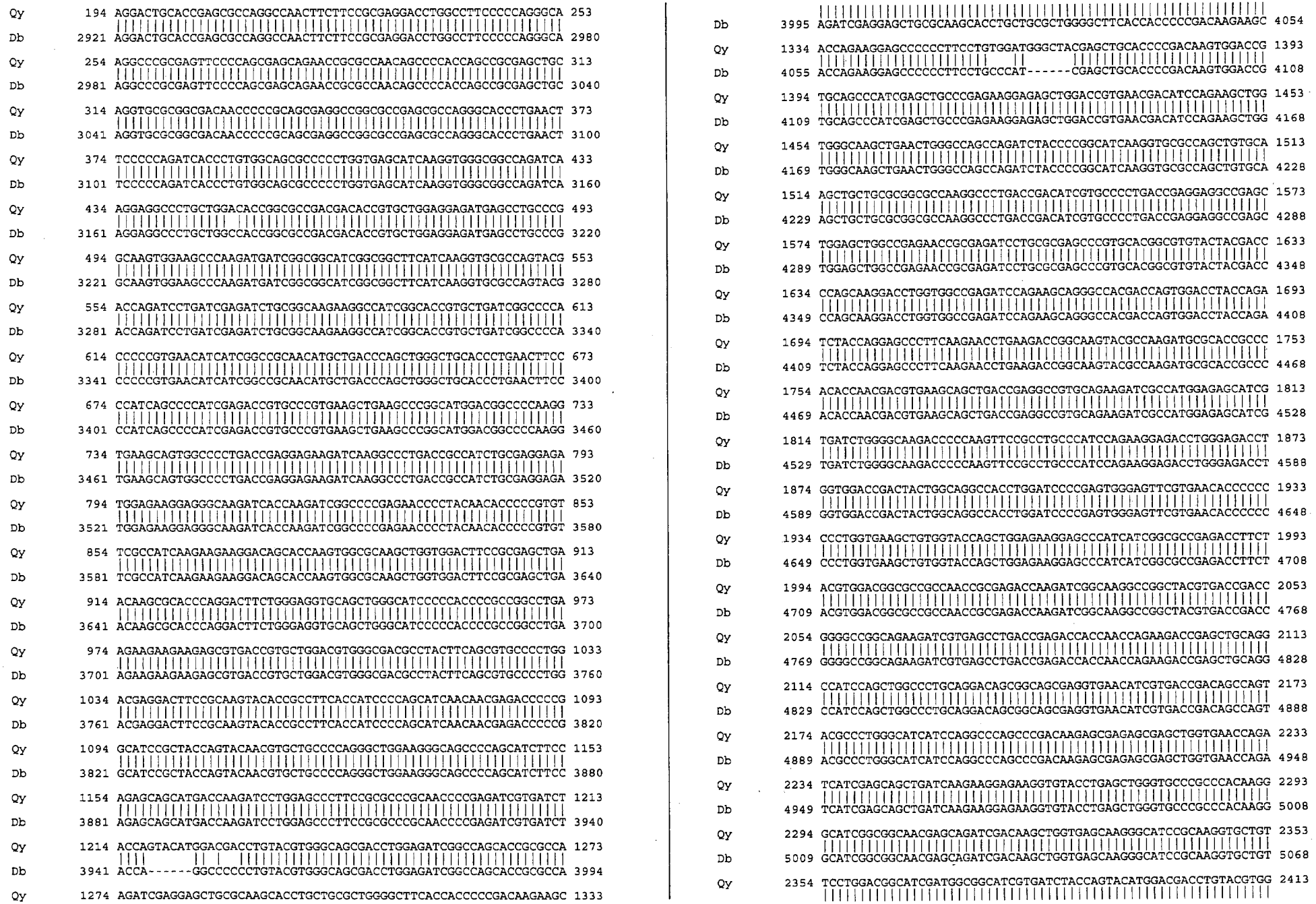
```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
```

```
Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
            ||||  ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy      1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy      1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
            ||||||||||||||||||||  ||      ||||||||||||||||||||||||||||||||
Db      2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy      1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy      1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy      1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy      1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy      1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy      1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy      1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy      1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy      1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy      1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy      1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy      2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy      2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy      2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy      2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy      2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy      2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy      2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 10
US-10-190-435-58
; Sequence 58, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 5184
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C
US-10-190-435-58

Query Match 96.9%; Score 2393.2; DB 14; Length 5184;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

```
Qy        14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy       134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2861 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920
```


Qy 194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db 2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy 1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db 3941 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 3994
Qy 1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db 3995 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 4054
Qy 1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db 4055 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 4108
Qy 1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db 4109 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 4168
Qy 1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db 4169 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 4228
Qy 1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db 4229 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 4288
Qy 1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db 4289 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 4348
Qy 1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db 4349 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 4408
Qy 1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db 4409 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 4468
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db 4469 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 4528
Qy 1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db 4529 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 4588
Qy 1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db 4589 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 4648
Qy 1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db 4649 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 4708
Qy 1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db 4709 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 4768
Qy 2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db 4769 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 4828
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db 4829 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 4888
Qy 2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db 4889 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 4948
Qy 2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
Db 4949 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 5008
Qy 2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db 5009 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 5068
Qy 2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413

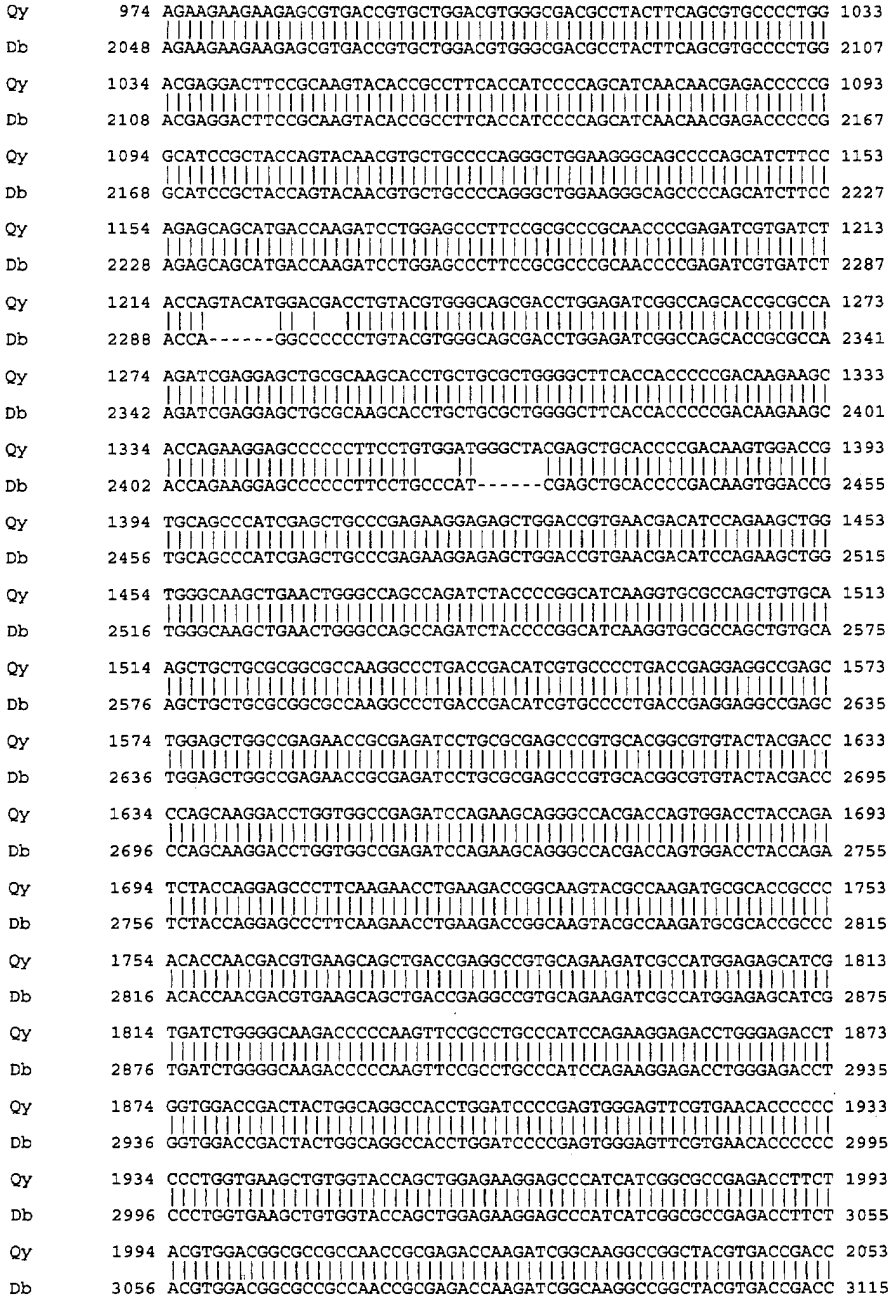
```
Db       5069 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 5128

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5129 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178


RESULT 11
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: TatRevNefgagCpolIna C
US-10-190-305A-82

  Query Match            96.9%;  Score 2393.2;  DB 14;  Length 5184;
  Best Local Similarity  99.2%;  Pred. No. 0;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy         614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy         674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy         734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy         794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy         854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy         914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy         974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy        1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy        1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy        1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy        1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
               ||||       || | ||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 3994

Qy        1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3995 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 4054

Qy        1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
               |||||||||||||||||||    ||      ||||||||||||||||||||||||||||
Db       4055 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 4108

Qy        1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4109 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 4168

Qy        1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4169 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 4228

Qy        1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4229 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 4288

Qy        1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4289 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 4348

Qy        1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4349 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 4408
```

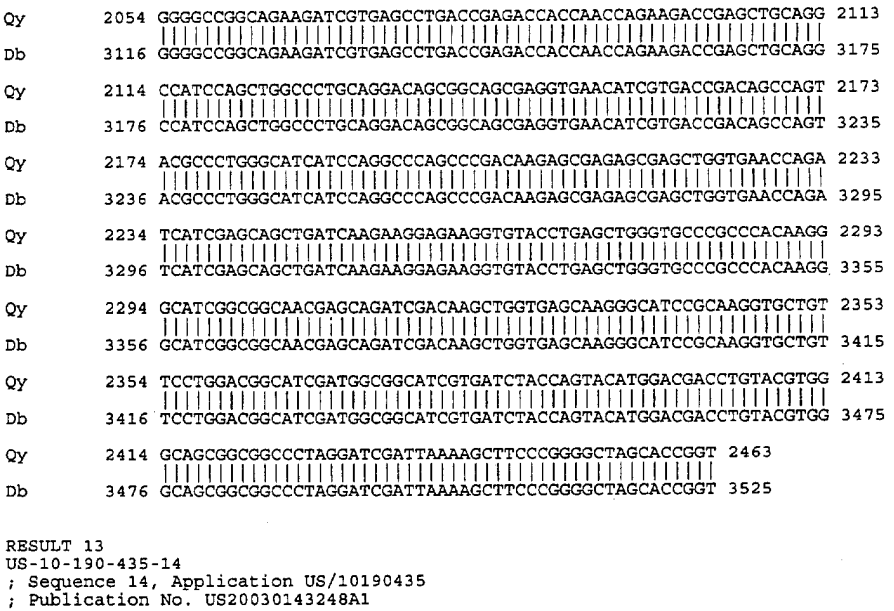
```
Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4409 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 4468

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4469 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 4528

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4529 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 4588

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4589 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 4648

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4649 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 4708

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4709 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 4768

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4769 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 4828

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACGAGCCAGT 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4829 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACGAGCCAGT 4888

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4889 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 4948

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4949 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 5008

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5009 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 5068

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5069 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 5128

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5129 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 12
US-10-190-435-13
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 3531
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13

Query Match 95.7%; Score 2362.8; DB 14; Length 3531;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2411; Conservative 0; Mismatches 27; Indels 12; Gaps 2;

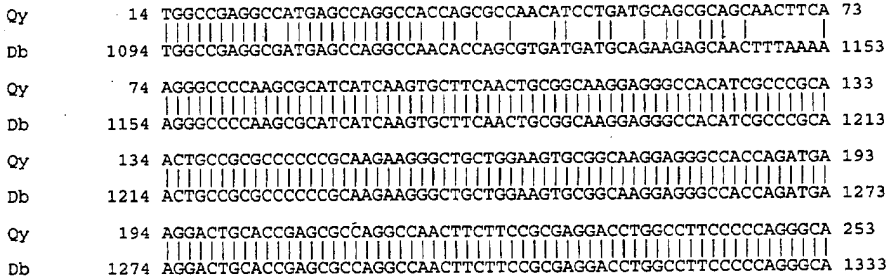
```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||  |||||||||||||  |  |  ||  ||  |  ||| |      |
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047
```

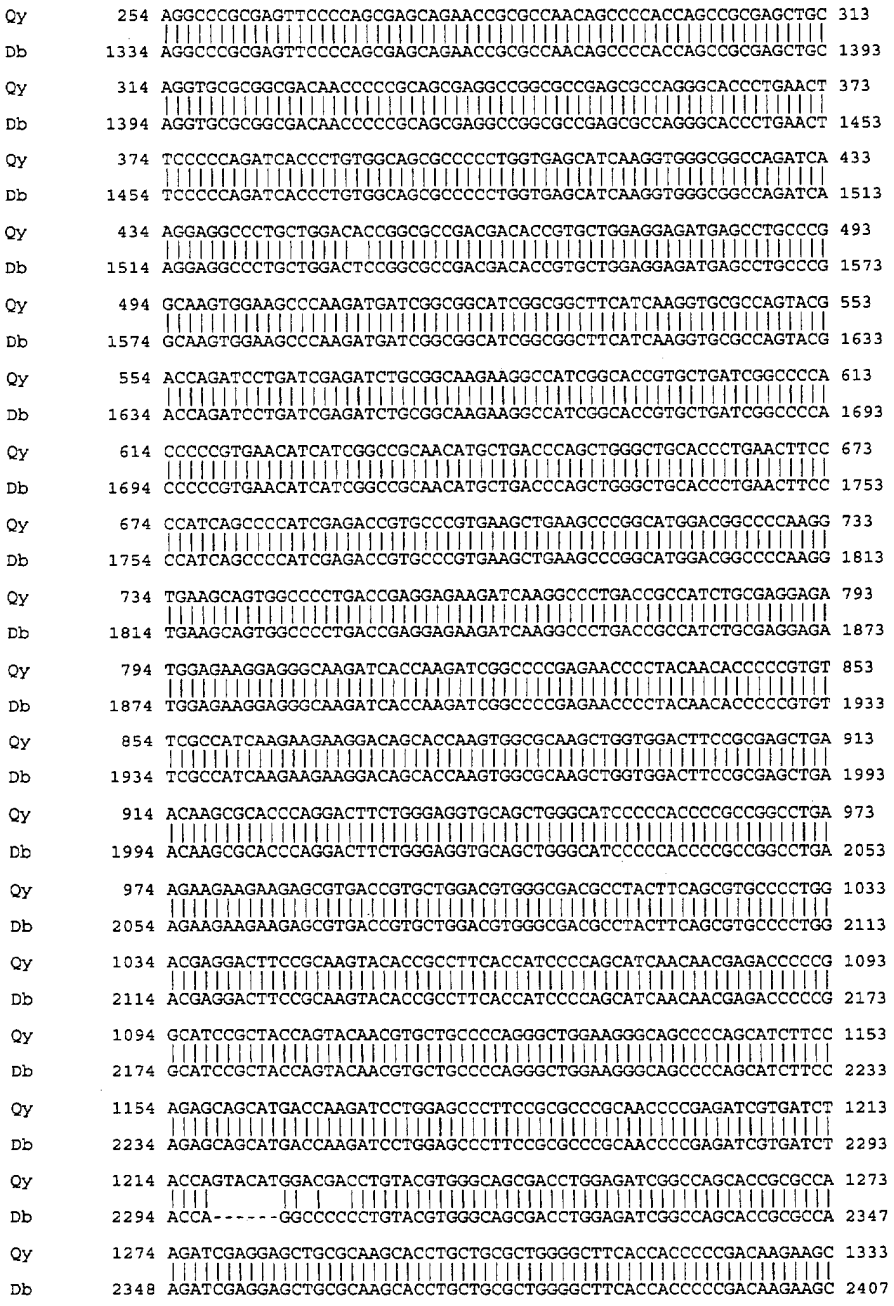




RESULT 13
US-10-190-435-14
; Sequence 14, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 3537
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmutAtt_C
US-10-190-435-14

Query Match 95.6%; Score 2361.2; DB 14; Length 3537;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2410; Conservative 0; Mismatches 28; Indels 12; Gaps 2;




Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 1334 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1393
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 1394 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1453
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 1454 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1513
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 1514 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1573
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 1574 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1633
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 1634 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1693
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 1694 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1753
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 1754 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1813
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 1814 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1873
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 1874 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1933
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 1934 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1993
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 1994 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2053
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 2054 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2113
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 2114 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2173
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 2174 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2233
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2234 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2293
Qy 1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db 2294 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2347
Qy 1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db 2348 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2407
Qy 1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db 2408 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2461
Qy 1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db 2462 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2521
Qy 1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db 2522 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2581
Qy 1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db 2582 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 2641
Qy 1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db 2642 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 2701
Qy 1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db 2702 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 2761
Qy 1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db 2762 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 2821
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db 2822 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 2881
Qy 1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db 2882 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 2941
Qy 1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db 2942 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3001
Qy 1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db 3002 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3061
Qy 1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db 3062 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3121
Qy 2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db 3122 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3181
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db 3182 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3241
Qy 2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db 3242 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3301
Qy 2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
Db 3302 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3361
Qy 2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db 3362 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3421
Qy 2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
Db 3422 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3481
Qy 2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3482 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3531
```

RESULT 14
US-10-190-435-15
; Sequence 15, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 3537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagPolmutIna_C
US-10-190-435-15

  Query Match   95.6%; Score 2361.2; DB 14; Length 3537;
  Best Local Similarity   98.4%; Pred. No. 0;
  Matches 2410; Conservative   0; Mismatches   28; Indels   12; Gaps   2;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| |||||||||||| || |  |   ||  ||  | || ||| |      |
Db       1094 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1153

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1154 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1213

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1214 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1273

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1274 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1333

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1334 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1393

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1394 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1453

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1454 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1513

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1514 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1573

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1574 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1633

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1634 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1693

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1694 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1753

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1754 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1813

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1814 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1873

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1874 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1933

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1934 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1993

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1994 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2053

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2054 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2113

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2114 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2173

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2174 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2233

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2234 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2293

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||    ||  |  ||||||||||||||||||||||||||||||||||||||||||||
Db       2294 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2347

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2348 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2407

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              ||||||||||||||||||||    ||      ||||||||||||||||||||||||||||
Db       2408 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2461

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2462 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2521

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2522 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2581

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2582 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 2641

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2642 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 2701

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2702 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 2761

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
```

```
Db       2762 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 2821

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813

Db       2822 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 2881

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873

Db       2882 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 2941

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933

Db       2942 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3001

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993

Db       3002 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3061

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053

Db       3062 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3121

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113

Db       3122 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3181

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173

Db       3182 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3241

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233

Db       3242 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3301

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293

Db       3302 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3361

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353

Db       3362 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3421

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413

Db       3422 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3481

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       3482 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3531


RESULT 15
US-10-190-435-12
; Sequence 12, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 5145
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: GagComplPolmutInaTatRevNef_C
US-10-190-435-12

  Query Match             95.6%;  Score 2360.2;  DB 14;  Length 5145;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2397;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGGCAGCAACTTCA 73

Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373

Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433

Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493

Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553

Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613

Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673

Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733

Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793

Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853

Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913

Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973

Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
```

```
Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              ||||||||||||||||||||  ||  ||||||||||||||||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGG 2430
              |||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGG 3891
```

Search completed: April 10, 2004, 21:24:32
Job time : 604.779 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 06:34:42 ; Search time 4127.92 Seconds
(without alignments)
17774.420 Million cell updates/sec

Title: US-09-610-313-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	90.2	3.7	2598	11	AY103647	AY103647 Zea mays
	2	85.6	3.5	951	12	BM321451	BM321451 rockefell
c	3	80.6	3.3	869	14	CK159167	CK159167 FGAS04056
	4	80.2	3.3	1132	12	BM320864	BM320864 rockefell
	5	79.6	3.2	1165	12	BM320900	BM320900 rockefell
	6	79	3.2	867	12	BM321430	BM321430 rockefell
	7	76.6	3.1	1550	12	BM321022	BM321022 rockefell
	8	75	3.1	545	12	BI724851	BI724851 1031075E0
	9	75	3.1	862	12	BM321023	BM321023 rockefell
	10	74.4	3.0	914	28	BZ568300	BZ568300 pacs2-164
	11	73.8	3.0	2433	29	AY401196	AY401196 Homo sapi
	12	73.2	3.0	853	12	BM321393	BM321393 rockefell
c	13	72.8	3.0	566	12	BM587428	BM587428 170006873
	14	72.6	3.0	788	14	CB643171	CB643171 OSJNEb03L
	15	72.4	2.9	753	29	CC675888	CC675888 OGWCO51TH
	16	72.2	2.9	640	10	BE601575	BE601575 HVSMEh009
	17	72.2	2.9	688	14	CB648640	CB648640 OSJNEb12C
	18	72.2	2.9	764	14	CB651670	CB651670 OSJNEb16L
	19	72.2	2.9	766	14	CB642928	CB642928 OSJNEb03F
	20	72.2	2.9	809	14	CB641397	CB641397 OSJNEb01A
	21	72.2	2.9	841	14	CB651502	CB651502 OSJNEb16H
c	22	72.2	2.9	851	29	CG260623	CG260623 OGWGE95TV
	23	71.8	2.9	1304	28	BZ577729	BZ577729 msh2_5546
	24	71.8	2.9	1933	11	AY325173	AY325173 Rattus no
	25	71.6	2.9	500	12	BM372120	BM372120 EBro03_SQ
	26	71.6	2.9	538	12	BM368580	BM368580 EBem08_SQ
	27	71.6	2.9	540	9	AJ471121	AJ471121 AJ471121
	28	71.6	2.9	566	13	BQ464692	BQ464692 HF02P20r
	29	71.6	2.9	575	13	BU976068	BU976068 HA03B08r
	30	71.6	2.9	576	13	BU984666	BU984666 HF04K03r
	31	71.6	2.9	578	12	BM377112	BM377112 EBem05_SQ
	32	71.6	2.9	579	10	BF253640	BF253640 HVSMEf000
	33	71.6	2.9	582	12	BM372871	BM372871 EBma04_SQ
	34	71.6	2.9	586	13	BU994273	BU994273 HM06H03r
	35	71.6	2.9	588	13	BQ765498	BQ765498 EBro03_SQ
	36	71.6	2.9	596	13	BU997745	BU997745 HI08O07r
	37	71.6	2.9	600	9	AL508778	AL508778 AL508778
	38	71.6	2.9	615	13	CA009718	CA009718 HU14O18r
	39	71.6	2.9	621	13	BU986998	BU986998 HF13I08r
	40	71.6	2.9	625	13	BQ768295	BQ768295 EBro08_SQ
	41	71.6	2.9	637	13	BU999333	BU999333 HI14D01r
	42	71.6	2.9	641	13	BU978712	BU978712 HA14B02r
	43	71.6	2.9	645	13	BU977527	BU977527 HA11J06r
	44	71.6	2.9	646	13	BU968956	BU968956 HB10B12r
	45	71.6	2.9	655	13	BU996444	BU996444 HM13K04r

ALIGNMENTS

RESULT 1
AY103647
LOCUS AY103647 2598 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2598)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2598)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES Location/Qualifiers

source 1. .2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

ORIGIN

Query Match 3.7%; Score 90.2; DB 11; Length 2598;
Best Local Similarity 42.2%; Pred. No. 0.24;
Matches 814; Conservative 0; Mismatches 1098; Indels 15; Gaps 5;

```
Qy       296 CCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGC 355
Db        51 CCACCACCTCCCACCTCCTGCCACCGCGCGGCCCCAACCAACCACACCGCAGCGCAGCAA 110

Qy       356 GCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCA 415
Db       111 TGGCGGGCAACGGCGCCATCGTGGAGAGCGACCCGCTGAACTGGGGCGCGGCGGCGGCGG 170

Qy       416 AGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGG 475
Db       171 AGCTGGCCGGGAGCCACCTGGACGAGGTGAAGCGCATGGTGGCGCAGGCCCGGCAGCCCG 230

Qy       476 AGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA 535
Db       231 TGGTCAAGATCGAGGGTTCCACCCTCCGCGTCGGCCAGGTGGCCGCCGTCGCCTCCGCCA 290

Qy       536 TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCA 595
Db       291 AGGACGCGTCCGGCGTCGCCGTCGAGCTCGACGAGGAGGCCCGCCCCCGCGTCAAGGCCA 350

Qy       596 CCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGG 655
Db       351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410

Qy       656 GCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCG 715
Db       411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470

Qy       716 GCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA 775
Db       471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530

Qy       776 CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACC 835
Db       531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590

Qy       836 CCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGG 895
Db       591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650

Qy       896 TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCC 955
Db       651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710

Qy       956 CCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT 1015
Db       711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767

Qy      1016 ACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCA 1075
```

```
Db       768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827

Qy      1076 TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGG 1135
Db       828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887

Qy      1136 GCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCA 1195
Db       888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCT---GTCCGCCGTCTTCTGCG 944

Qy      1196 ACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db       945 AGGTCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACC 1004

Qy      1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG---CGCTGGGGCTTCACCA 1312
Db       1005 CGGGGTCCATCGAGGCCGCGGCCATCATGGAGCACATCCTGGATGGCAGCTCCTTCATGA 1064

Qy      1313 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACA 1372
Db       1065 AGCAGGCCAAGAAGTTGAACGAGCTGGACCCGCTGCTGAAGCCCAAGCAGGACAGGTACG 1124

Qy      1373 AGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCC 1432
Db       1125 CGCTCCGCACGTCGCCGCAGTGGCTGGGCCCCCAGATCGAGGTCATCCGCGCCGCCACCA 1184

Qy      1433 AGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCC 1492
Db       1185 AGTCCATCGAGCGCGAGGTCAACTCCGTGAACGA---CAACCCGGTCATCGACGTCCACC 1241

Qy      1493 AGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGG 1552
Db       1242 GCGGCAAGGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTGTCCATGGACA 1301

Qy      1553 AGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGT 1612
Db       1302 ACGCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGAGCTCG 1361

Qy      1613 ACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGA 1672
Db       1362 TCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGCAACCCCAGCC 1421

Qy      1673 CCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGC 1732
Db       1422 TGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCCGAGCTCCAGT 1481

Qy      1733 GCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGG 1792
Db       1482 ACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGAGCAGCACAACCAGGACGTGA 1541

Qy      1793 AGAGCAT---CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849
Db       1542 ACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGGCGATCGACATCCTGAAGCTCA 1601

Qy      1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909
Db       1602 TGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCACCTCGAGGAGA 1661

Qy      1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969
Db       1662 ACATCAAGGCGTCGGTGAAGAACACCGTGACCCAGGTGGCCAAGAAGGTGCTGACCATGA 1721

Qy      1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
Db       1722 ACCCCTCGGGCGAGCTCTCCAGCGCCCGCTTCAGCGAGAAGGAGCTGATCAGCGCCATCG 1781

Qy      2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089
Db       1782 ACCGCGAGGCCGTGTTCACGTACGCGGAGGACGCGGCCAGCGGCAGCCTGCCGCTGATGC 1841

Qy      2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149
Db       1842 AGAAGCTGCGCGCCGTGCTGGTGGACCACGCCCTCAGCAGCGGCGACGCGGAGCGGGAGC 1901
```

```
Qy       2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209
Db       1902 CCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGCTGCCCCAGG 1961

Qy       2210 TGGTGAA 2216
Db       1962 AGGTGGA 1968
```

RESULT 2
BM321451
LOCUS BM321451 951 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.
ACCESSION BM321451
VERSION BM321451.1 GI:18055857
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 951)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 951 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .951
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN

Query Match 3.5%; Score 85.6; DB 12; Length 951;
Best Local Similarity 46.0%; Pred. No. 0.69;
Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

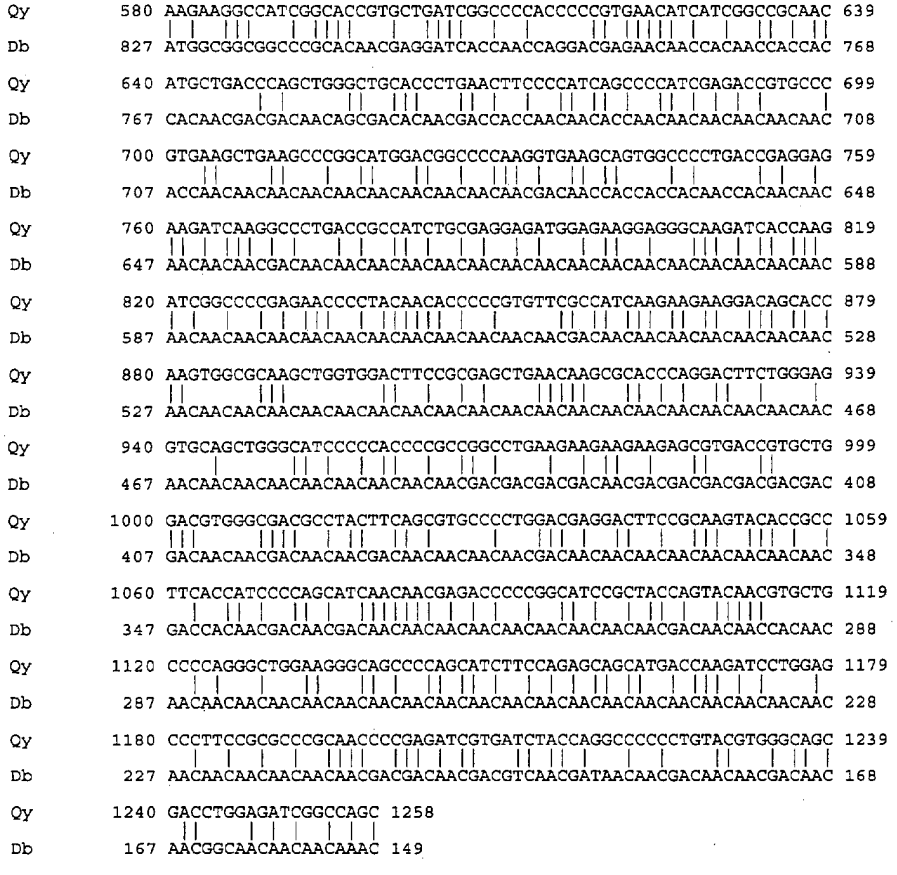
```
Qy       1722 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1781
Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1782 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1841
Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1842 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1901
Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1902 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1961
Db        399 CGACGGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1962 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2021
Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2022 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2081
Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2082 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2141
Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2142 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2201
Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy       2202 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2261
Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2262 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2321
Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2322 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2382 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2423
Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917
```

RESULT 3
CK159167/c
LOCUS CK159167 869 bp mRNA linear EST 05-DEC-2003
DEFINITION FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK159167
VERSION CK159167.1 GI:38985053
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 869)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_ests@cs.usask.ca
This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [128,636].
Plate: TaLt537 row: N column: 23.
FEATURES Location/Qualifiers
source 1. .869
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaLt5"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days

(equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"

ORIGIN

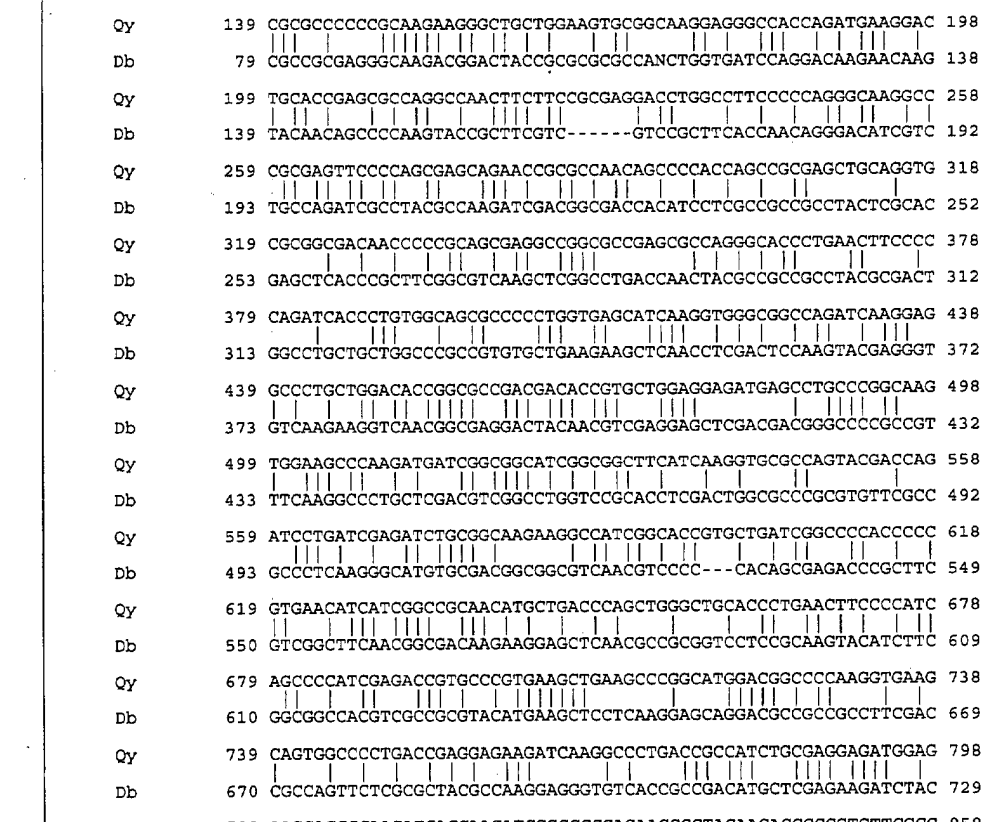
Query Match 3.3%; Score 80.6; DB 14; Length 869;
Best Local Similarity 44.9%; Pred. No. 2.7;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;



RESULT 4
BM320864
LOCUS BM320864 1132 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.
ACCESSION BM320864
VERSION BM320864.1 GI:18055270
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1132 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1132
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.3%; Score 80.2; DB 12; Length 1132;
Best Local Similarity 45.3%; Pred. No. 3.1;
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;



```
Db        730 ACCGAGGCCCACAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAG 789

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918

Db        790 CCCGAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966

Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897


RESULT 5
BM320900
LOCUS       BM320900                1165 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320900
VERSION     BM320900.1  GI:18055306
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1165   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1165
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79.6;  DB 12;  Length 1165;
  Best Local Similarity   45.1%;  Pred. No. 3.7;
  Matches  428;  Conservative    0;  Mismatches  506;  Indels   16;  Gaps    3;

Qy         90 CATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCG 149

Db         46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTCCAGACCCAGTTCCGTCGCCGCCGCGAGGG 105

Qy        150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209

Db        106 CAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAGTACAACAGCCC 165

Qy        210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269

Db        166 CAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTCTGCCAGATCGC 219

Qy        270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329

Db        220 CTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCACGAGCTCACCCG 279

Qy        330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389

Db        280 CTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACTGGCCTGCTGCT 339

Qy        390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGA 449

Db        340 GGCCCGCCGTGTGCTGAAGAAGCTCAACCTCTACTCCAAGTACGAGGGTGTCAAGAAGGT 399

Qy        450 CACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAA 509

Db        400 CAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGTTCAAGGCCC-- 457

Qy        510 GATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGA 569

Db        458 --TGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCCGCCCTCAA 515

Qy        570 GATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCAT 629

Db        516 GGGCATGTGCGACGGCGGCGTCAACGTCCCCCACAGCGAGACCCGCTTCGTCGGCTTCAA 575

Qy        630 CGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGA 689

Db        576 CGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCACGT 635

Qy        690 GACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCT 749

Db        636 CGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGACCGCCAGTTCTC 695

Qy        750 GACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAA 809

Db        696 GCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTACACCGAGGCCCA 755

Qy        810 GATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAA 869

Db        756 ------CAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAGCCCGA 809

Qy        870 GGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGA 929

Db        810 GGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAGAACCG 869

Qy        930 CTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGT 989

Db        870 CGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCCAAGAGCCAGTAAATTCC 929

Qy        990 GACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGG 1039

Db        930 CGGTGCCCTGTACACCGCCATCTCCCGCTGCGGCGTTGCGCTGCTNGCGG 979


RESULT 6
BM321430
LOCUS       BM321430                867 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
            sequence.
ACCESSION   BM321430
VERSION     BM321430.1  GI:18055836
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 867   Std Error: 0.00
```

```
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79;  DB 12;  Length 867;
  Best Local Similarity   45.7%;  Pred. No. 4.1;
  Matches  314;  Conservative    0;  Mismatches  370;  Indels    3;  Gaps    1;

Qy         58 CAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAG 117
Db        131 CCGCACAAGATGCGCGAGTGCCTGCCGGTCATCATCCTGCTGCGCAACAGGTTGAAGTAC 190

Qy        118 GGCCACATCGCCCGCAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAG 177
Db        191 GCGCTGACCCGCCGTGAGGTCACCTCGATCGTGATGCAGCGCCTGATCAAGATCGACGGC 250

Qy        178 GAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237
Db        251 AAGGTCCGCACCGACACGCACTTCCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310

Qy        238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCC 297
Db        311 ACCGACGAGCACTTCCGCCTGCTCTACGACACCAAGGGCCGCTTCCAGGCGCACCGCATC 370

Qy        298 ACCAGCCGCGAGCTGCAGGT---GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354
Db        371 AACTCGACCGAGGCCAAGTTCAAGCTCGGCAAGGTCCGCCGCGTGCAGCTCGGCAACAAG 430

Qy        355 CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATC 414
Db        431 GGCATCCCGTACCTGGTGACCCACGACGCGCGCACGATCCGCTACCCCAACCCCGACATC 490

Qy        415 AAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTG 474
Db        491 AAGGTCAACGACACGGTCAAGATCGACCTGCGCTCGGGCAAGATCATCGACTTCGTCAGG 550

Qy        475 GAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTC 534
Db        551 TTCGAGATCGGCAACCTCGTCATGATCACTGGCGGACGCAACCTTGGCCGCGTCGGCGTC 610

Qy        535 ATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGC 594
Db        611 ATTGTGCGCCGCGAGAAGCACGAGGGCTCGTTCGAGATCATCCACGTCAAGGACGCCGTC 670

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654
Db        671 GGCCACCAGTTCGCGACGCGCCTGACCAACGTCTTCGTGATCGGCAAGGGCACCAAGTCC 730

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714
Db        731 CTCGTCACGCTGCCCGCCGGCAAGGGCATCAAGAAGTCGATCATCGAGGAGTTCCAGGCG 790

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAG 741
Db        791 CGCCACGCGCACAGGACCGAGCAGGAG 817

RESULT 7
BM321022
LOCUS       BM321022                1550 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1192 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321022
VERSION     BM321022.1  GI:18055428
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1550)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1550   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1550
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.1%;  Score 76.6;  DB 12;  Length 1550;
  Best Local Similarity   45.6%;  Pred. No. 9;
  Matches  308;  Conservative    0;  Mismatches  364;  Indels    3;  Gaps    1;

Qy       1749 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1808
Db          5 CGGCACGAGGACGCAGGACCACGCGGCCGCCGCCATCGCGCAGGCCGGCGTCTCGGTCTT 64

Qy       1809 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1868
Db         65 CGCCTGGAAGGGCGAGAACCTCCAGGAGTACTGGGAGTGCACCTGGAAGGCCCTGTGCTT 124

Qy       1869 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1928
Db        125 CGGCCCCTACCAGGGCCCTCAGATCATCGTCGACGACGGCGGTGACGCGACTCTGATGAT 184

Qy       1929 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1988
Db        185 CCACAAGGGGTTCGCGGCCGAGGACAACCCCAAGCTGCTGGAGGACGACGAGGGCCTCGA 244

Qy       1989 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2048
Db        245 GGAGGTCGCCTGCCTCAACAACGTGCTCAAGCAGGTCCAGAAGGAGCAGCCCGGCTTCTG 304

Qy       2049 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2108
Db        305 GCACAAGATCCTCCCCGAGATCCGCGGTGTCAGCGAGGAGACGACGACTGGCGTGATGAG 364

Qy       2109 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2168
Db        365 GCTGTACCAGCTGCACCGCGACGGCAAGCTGCTGTTCCCGGCCGTCAACGTCAACGACTC 424

Qy       2169 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2228
Db        425 TG---TCACCAAGAGCAAGTTTGACAACATCTACGGCTGCCGCCACTCGCTCATCGACGG 481

Qy       2229 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2288
Db        482 CATCAAGCGCGGCACCGACGTGATGCTCGGCGGCAAGGTCGCCGTCGTCGCGGGCTACGG 541

Qy       2289 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2348
Db        542 CGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCGCGGCCAGGGCTGCCGCGTCATCGTGAC 601
```

```
Qy       2349 CGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGG 2408
Db        602 GGAGATCGACCCCATCTGCGCGCTGCAGGCGTCGATGGCGGGCTTCGAGGTCAACACGCT 661

Qy       2409 CGGCCCTAGGATCGA 2423
Db        662 CGAGGCGGGGCTCGA 676
```

RESULT 8
BI724851
LOCUS BI724851 545 bp mRNA linear EST 19-SEP-2001
DEFINITION 1031075E05.y1 C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI724851
VERSION BI724851.1 GI:15700546
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 545)
AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .545
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoRI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."
ORIGIN

Query Match 3.1%; Score 75; DB 12; Length 545;
Best Local Similarity 46.7%; Pred. No. 11;
Matches 237; Conservative 0; Mismatches 270; Indels 0; Gaps 0;

```
Qy       1184 TCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACC 1243
Db         36 TGCGCGCGCAGCAGCTGTGGTTTGCCATCCTGGAGGCCCAGGTGGAGACCGGCAACCCCT 95

Qy       1244 TGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGG 1303
Db         96 ACATGCTGTACAAGGACCACTGCAACCGCAAGTCCAACCAGCAGAACCTGGGCACCATCA 155

Qy       1304 GCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGC 1363
Db        156 AGTGCTCCAACCTGTGCACGGAGATTATCGAGTACACCAGCCCCGATGAGATCGCCGTGT 215

Qy       1364 ACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGA 1423
Db        216 GCAACCTGGCCTCCATCGCGCTGCCCCGCTTCGTGCGCGAGCGCGGCAGCCAGCCCGGCT 275

Qy       1424 ACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCA 1483
Db        276 GGGAGAGCAAGAAGCTGGTGGGCTCGCTGGACGCCGCCAACCGCTTCTTCGACTTCGAGA 335

Qy       1484 AGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCC 1543
Db        336 AGCTGGCTGAGATGACCAAGATGGTGACCCGCAACCTGAACAAGATCATTGACGTCAACT 395

Qy       1544 TGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGC 1603
Db        396 ACTACCCGCACCCCAAGGCGCAGCGCTCCAACATGCGCCACCGCCCCATCGGCCTGGGCG 455

Qy       1604 ACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACG 1663
Db        456 TGCAGGGCCTGGCGGACACCTTCATCCTGCTGGGCATGGCCTTCGACTCGCCCGAGGCCG 515

Qy       1664 ACCAGTGGACCTACCAGATCTACCAGG 1690
Db        516 CCGAGCTGAACCGCCAGATCTTCGAGG 542
```

RESULT 9
BM321023
LOCUS BM321023 862 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.594 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.
ACCESSION BM321023
VERSION BM321023.1 GI:18055429
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 862)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 862 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .862
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN

Query Match 3.1%; Score 75; DB 12; Length 862;
Best Local Similarity 46.2%; Pred. No. 12;
Matches 283; Conservative 0; Mismatches 327; Indels 3; Gaps 1;

```
Qy       1722 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1781
```

```
Db        220 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 279

Qy       1782 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1841

Db        280 CATCGCGCAGGCCGGAGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 339

Qy       1842 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1901

Db        340 GGAGTGCACCTNGAAGGCCCTGTGCTTCGGCCCCTACCAGGGACCTCAGANCATCGTCGA 399

Qy       1902 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1961

Db        400 CGACGGCGGTGACGCGACTCTAATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 459

Qy       1962 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2021

Db        460 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 519

Qy       2022 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2081

Db        520 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 579

Qy       2082 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2141

Db        580 CGAGGAGACGACGACTGGCGTGATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 639

Qy       2142 CATCGTGACCGACAGCCAGTACGCCCTGGGCATC---ATCCAGGCCCAGCCCGACAAGAG 2198

Db        640 GTTCCCGGCCGTCAACGTCAACGACTCTGTCACCAAGAGCAAGTTTGACAACATCTACGG 699

Qy       2199 CGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCT 2258

Db        700 CTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGGCAA 759

Qy       2259 GAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAG 2318

Db        760 GGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCGCGG 819

Qy       2319 CAAGGGCATCCGC 2331

Db        820 CCAGGGCTGCCGC 832
```

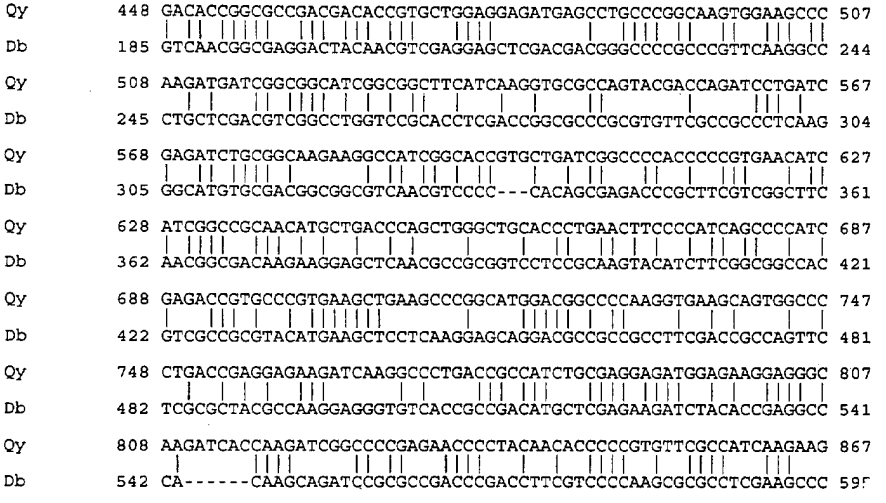
```
RESULT 10
BZ568300
LOCUS       BZ568300                914 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_7463.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_7463, genomic survey sequence.
ACCESSION   BZ568300
VERSION     BZ568300.1  GI:27201058
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .914
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_7463"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN
```

Query Match 3.0%; Score 74.4; DB 28; Length 914;
Best Local Similarity 47.0%; Pred. No. 15;
Matches 265; Conservative 0; Mismatches 296; Indels 3; Gaps 1;

```
Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711

Db        112 CTGCGCAACCAGCAGGACGGCCTGATCTTCAGCCAGGCCGGCGCCCTCGAGGTGCAGGCC 171

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771

Db        172 GGCAGCCTGGACAACCGCCAGGGCACGCTCCAGGCCCAGGGCGACAACCAGCTGCGTATC 231

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831

Db        232 GGTGGCGCGCTGGACAACCAGGGCGGCCGCCTGGACAGCCGGGCCGGTAACCTAGACCTG 291

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891

Db        292 CAGAGCGGCAGCCTCGACAACGGCGCCGGCGGCGTGCTCAACAGCGCCAAGGGTTGGCTG 351

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951

Db        352 AAGCTGGTCACCGGGCTGTTCGACAACAGCGCCGGCGTCACCCAGGCGCAGTCGCTGGAA 411

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011

Db        412 ATCCGCGCCGGGCAAGGCGTGCGCAACCAGCAGGGC---CATGTCTCGGCGCTGGGCGGC 468

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071

Db        469 GACAACCGCATCGTCACCGTCGACTTCGACAACCAGGGTGGCGGCCTCTACGCCAGCGGC 528

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131

Db        529 CTGCTCAGCCTCGACGGCCAGCGCTTCCTCAACCAGGGCGCGGCGGCGGGCCAGGGCGGC 588

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191

Db        589 AAGGTCGGCGCCGGGCGCATCGACTTCAGCCTGGCCGGCGCGCTGGCCAACCGCTTCGCC 648

Qy       1192 CGCAACCCCGAGATCGTGATCTAC 1215

Db        649 AGTTGGAAAGCGAGAGCGAGCTGC 672
```

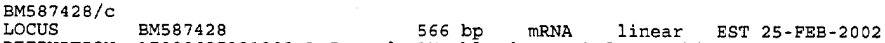
```
RESULT 11
AY401196
LOCUS       AY401196                2433 bp    DNA     linear   GSS 12-DEC-2003
DEFINITION  Homo sapiens HCM0810 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY401196
VERSION     AY401196.1  GI:39757185
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2433)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
```

```
   PUBMED   14671302
REFERENCE   2  (bases 1 to 2433)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1. .2433
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1. .>2433
                     /locus_tag="HCM0810"
ORIGIN

  Query Match             3.0%;  Score 73.8;  DB 29;  Length 2433;
  Best Local Similarity   43.0%;  Pred. No. 21;
  Matches  485;  Conservative    0;  Mismatches  632;  Indels   12;  Gaps    2;

Qy        463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db       1240 GACAAGCTCCGGCAGCGGCTCACCAAGAACGCGCAGGACAAGCTGGAGCTGCACCTGTTC 1299

Qy        523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db       1300 ATGCTCAGTGGCATCCCTGACACTGTGTTTGACCTGGTGGAGCTGGAGGTCCTCAAGCTG 1359

Qy        583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db       1360 GAGCTGATCCCCGACGTGACCATCCCGCCCAGCATTGCCCAGCTCACGGGCCTCAAGGAG 1419

Qy        643 CTG-------ACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG-----CCCCATCGAG 690
Db       1420 CTGTGGCTCTACCACACAGCGGCCAAGATTGAAGCGCCCGCGCTGGCCTTCCTGCGCGAG 1479

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db       1480 AACCTGCGGGCGCTGCACATCAAGTTCACCGACATCAAGGAGATCCCGCTGTGGATCTAT 1539

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db       1540 AGCCTGAAGACACTGGAGGAGCTGCACCTGACGGGCAACCTGAGCGCGGAGAACAACCGC 1599

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db       1600 TACATCGTCATCGACGGGCTGCGGGAGCTCAAACGCCTCAAGGTGCTGCGGCTCAAGAGC 1659

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db       1660 AACCTAAGCAAGCTGCCACAGGTGGTCACAGATGTGGGCGTGCACCTGCAGAAGCTGTCC 1719

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db       1720 ATCAACAATGAGGGCACCAAGCTCATCGTCCTCAACAGCCTCAAGAAGATGGCGAACCTG 1779

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db       1780 ACTGAGCTGGAGCTGATCCGCTGTGACCTGGAGCGCATCCCCCACTCCATCTTCAGCCTC 1839

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db       1840 CACAACCTGCAGGAGATTGACCTCAAGGACAACAACCTCAAGACCATCGAGGAGATCATC 1899

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db       1900 AGCTTCCAGCACCTGCACCGCCTCACCTGCCTTAAGCTGTGGTACAACCACATCGCCTAC 1959

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTAC 1230
Db       1960 ATCCCCATCCAGATCGGCAACCTCACCAACCTGGAGCGCCTCTACCTGAACCGCAACAAG 2019

Qy       1231 GTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCAC 1290
Db       2020 ATCGAGAAGATCCCCACCCAGCTCTTCTACTGCCGCAAGCTGCGCTACCTGGACCTCAGC 2079

Qy       1291 CTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG 1350
Db       2080 CACAACAACCTGACCTTCCTCCCTGCCGACATCGGCCTCCTGCAGAACCTCCAGAACCTA 2139

Qy       1351 CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1410
Db       2140 GCCATCACGGCCAACCGGATCGAGACGCTCCCTCCGGAGCTCTTCCAGTGCCGGAAGCTG 2199

Qy       1411 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1470
Db       2200 CGGGCCCTGCACCTGGGCAACAACGTGCTGCAGTCACTGCCCTCCAGGGTGGGCGAGCTG 2259

Qy       1471 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1530
Db       2260 ACCAACCTGACGCAGATCGAGCTGCGGGGCAACCGGCTGGAGTGCCTGCCTGTGGAGCTG 2319

Qy       1531 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACC 1579
Db       2320 GGCGAGTGCCCACTGCTCAAGCGCAGCGGCTTGGTGGTGGAGGAGGACC 2368


RESULT 12
BM321393
LOCUS       BM321393                 853 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1222 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM321393
VERSION     BM321393.1  GI:18055799
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 853   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.0%;  Score 73.2;  DB 12;  Length 853;
  Best Local Similarity   46.9%;  Pred. No. 20;
  Matches  299;  Conservative    0;  Mismatches  329;  Indels    9;  Gaps    2;

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db        125 CTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGTGTCAAGAAG 184
```



RESULT 13
BM587428/c
LOCUS BM587428 566 bp mRNA linear EST 25-FEB-2002
DEFINITION 17000687321202 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone 19600449696110 5', mRNA sequence.
ACCESSION BM587428
VERSION BM587428.1 GI:18883289
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE 1 (bases 1 to 566)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 w. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AAX row: G column: 08
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .566
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone="19600449696110"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cDNA.blood1"
/note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"
ORIGIN

Query Match 3.0%; Score 72.8; DB 12; Length 566;
Best Local Similarity 48.3%; Pred. No. 21;
Matches 203; Conservative 0; Mismatches 217; Indels 0; Gaps 0;



RESULT 14
CB643171
LOCUS CB643171 788 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNEb03L13.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA clone OSJNEb03L13 5', mRNA sequence.
ACCESSION CB643171
VERSION CB643171.1 GI:29638162
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 788)
AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288

```
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 03  row: L  column: 13
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES              Location/Qualifiers
     source           1. .788
                      /organism="Oryza sativa (japonica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:39947"
                      /clone="OSJNEb03L13"
                      /tissue_type="Leaf"
                      /dev_stage="3 week"
                      /lab_host="DH10B"
                      /clone_lib="OSJNEb"
                      /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                      XhoI; 24 hrs after innoculation with Rice Blast (Che
                      86061)"
ORIGIN

  Query Match             3.0%;  Score 72.6;  DB 14;  Length 788;
  Best Local Similarity   46.3%;  Pred. No. 24;
  Matches  314;  Conservative    0;  Mismatches  355;  Indels    9;  Gaps    2;

Qy        107 GCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGA 166
Db         99 GCGCGGCGGCGGCCGAGATGGCCGGCAGCCACCTCGACGAGGTGAAGCGCATGGTGGCGC 158

Qy        167 AGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCC 226
Db        159 AGTTCCGCGAGCCGCTGGTCAAGATCCAGGGCGCCACCCTCCGCGTCGGCCAGGTGGCCG 218

Qy        227 GCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCG 286
Db        219 CCGTCGCCCAGGCCAAGGACGCCGCCGGCGTCGCCGTCGAGCTCGACGAGGAGGCCCGCC 278

Qy        287 CCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCG 346
Db        279 CCCGCGTCAAGGCCAGCAGCGAGTGGATCCTCAACTGCATCGCCCACGGCGGCGACATCT 338

Qy        347 GCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGG 406
Db        339 ACGGCGTCACCACCGGCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGCCCCGCCC 398

Qy        407 TGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGA---CG 463
Db        399 TCCAAGTCGAGCTCCTCAGGCATCTCAACGCCGGAATCTTCGGCACTGGCTCCGATGGCC 458

Qy        464 ACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCA 523
Db        459 ACACGCTGCCGTCGGAGACGGTGCGGGCGGCCATGCTCGTGCGCATCAACACCCTCCTCC 518

Qy        524 TCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGA 583
Db        519 AGGGCTACTCCGGCATCCGGTTCGAGATCCTCGAGGCCATCACCAAGCTGCTCAACACCG 578

Qy        584 AGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGC 643
Db        579 GCGTCACGCCGTGCCTGCCGCTCCGTGGCACCATCACCGCGTCCGGTGACCTGGTTCCCC 638

Qy        644 TGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGA 703
Db        639 TGTCCTACATTGCCGGCCTCATCACCGGCCGCCCCAACGCGCAGGCCATCTCGCCCGACG 698

Qy        704 AGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGA 763
Db        699 GCAGGAAG------GTGGACGCCGCCGAGGCGTTNCAGCTCGCCGGCATCGAGGGTGGCT 752

Qy        764 TCAAGGCCCTGACCGCCA 781
Db        753 TCTTCACGCTGAACCCCA 770
```

```
RESULT 15
CC675888
LOCUS       CC675888                753 bp    DNA     linear   GSS 19-JUN-2003
DEFINITION  OGWCO51TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0539J05,
            genomic survey sequence.
ACCESSION   CC675888
VERSION     CC675888.1  GI:32080584
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGWCO51TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES              Location/Qualifiers
     source           1. .753
                      /organism="Zea mays"
                      /mol_type="genomic DNA"
                      /strain="B73"
                      /db_xref="taxon:4577"
                      /clone="ZMMBMa0539J05"
                      /clone_lib="ZM_0.7_1.5_KB"
                      /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                      methylation filtered genomic DNA library"
ORIGIN

  Query Match             2.9%;  Score 72.4;  DB 29;  Length 753;
  Best Local Similarity   46.2%;  Pred. No. 25;
  Matches  281;  Conservative    0;  Mismatches  321;  Indels    6;  Gaps    1;

Qy         31 CAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATC 90
Db         40 CACACCACCGACGACACCTTCCGCGGGGTCCGCGTCAAGTGGACCTCCACCACGCGCACC 99

Qy         91 ATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGC 150
Db        100 GTCGACCGCAGCGGCAGCGGCGGCGGCGGTAACCCGTACAACATATTCGGCCGCGGCGGC 159

Qy        151 AAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGC 210
Db        160 GACCAGCGCGGCCTCGAGCTCCAGTTCCCGCGCCAGCACCGGGACCTCGTCCACCACCAC 219

Qy        211 CAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCC 270
Db        220 TACATCCCGCACCTCATCGACGAGGCCACCCGGATGCGGCTCAAGTCGCGGGAGCGCCGC 279

Qy        271 AGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAAC 330
Db        280 CTCTACACCAACCGCGCCACGGGGCCCTGCGACGACCACCACCGCCTCTGGACGTCGCAC 339

Qy        331 CCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTG 390
Db        340 GCCTTCGCGCACCCGTCCACGTTCGACACGCTGGCGCTCGACCCGGCGCTGCGCGACGAG 399

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db        400 GTCCGCGCCGACCTGCTGCGCTTCGCCGCGCGCCGGGACCACTACGCGCGCGTCGGCCGC 459
```

```
Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              || |         |    | ||||  | |       ||| |||||||| | | ||| |||
Db        460 GCCTGGAAGCGGGGGTACCTGCTCCACG------GCCCGCCCGGCACGGGCAAGACCAGC 513

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              |  |||  | |||||  |  | |  || |  |   |  |||||||  |   || | |  |
Db        514 CTCGTCGCTGCCATCGCCAACCTGCTCGACTTCGACGTCTACGACCTGGAGCTCACCACG 573

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |   | |||       | || ||  | ||||  ||  | |||| ||| |  | || ||
Db        574 GTGCCCACCAACTCCCACCTCCGCCGCCTGCTCGTCTCCACCACGCCCAAGTCCGTCGTC 633

Qy        631 GGCCGCAA 638
              | | | |
Db        634 GTCGTCGA 641
```

Search completed: April 10, 2004, 16:16:27
Job time : 4130.92 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 10, 2004, 02:45:46 ; Search time 619.487 Seconds
(without alignments)
16849.133 Million cell updates/sec

Title: US-09-610-313-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	100.0	2457	6	ABL39961	Abl39961 Synthetic
2	2443.4	99.4	2445	7	ACA03546	Aca03546 Synthetic
3	2443.4	99.4	2445	9	ADC13264	Adc13264 DNA of HI
4	2436.4	99.2	3930	9	ADC13230	Adc13230 DNA of HI
5	2436.2	99.2	2463	6	ABL39960	Abl39960 Synthetic
6	2434.8	99.1	3930	9	ADC13231	Adc13231 DNA of HI
7	2434.8	99.1	3930	9	ADC13232	Adc13232 DNA of HI
8	2434.8	99.1	5184	7	ACA03591	Aca03591 Synthetic
9	2434.8	99.1	5184	9	ADC13279	Adc13279 DNA of HI
10	2428.6	98.8	2457	7	ACA03547	Aca03547 Synthetic
11	2428.6	98.8	2457	9	ADC13265	Adc13265 DNA of HI
12	2415.4	98.3	2469	6	ABL39959	Abl39959 Synthetic
13	2404.4	97.9	3531	9	ADC13234	Adc13234 DNA of HI
14	2403.4	97.8	2457	7	ACA03548	Aca03548 Synthetic
15	2403.4	97.8	2457	9	ADC13266	Adc13266 DNA of HI
16	2402.8	97.8	3537	9	ADC13236	Adc13236 DNA of HI
17	2401.8	97.8	5145	7	ACA03521	Aca03521 Synthetic
18	2401.8	97.8	5145	9	ADC13233	Adc13233 DNA of HI
19	2391.8	97.3	3538	9	ADC13235	Adc13235 DNA of HI
20	2343.4	95.4	3607	7	ACA03551	Aca03551 Synthetic
21	2343.4	95.4	3607	9	ADC13269	Adc13269 DNA of HI
22	2325.2	94.6	3597	7	ACA03549	Aca03549 Synthetic
23	2325.2	94.6	3597	9	ADC13267	Adc13267 DNA of HI
24	2295.8	93.4	3624	7	ACA03550	Aca03550 Synthetic
25	2295.8	93.4	3624	9	ADC13268	Adc13268 DNA of HI
26	2136	86.9	2460	7	ACA03541	Aca03541 Synthetic
27	2136	86.9	2460	7	ACC78505	Acc78505 HIV p2Pol
28	2135	86.9	3564	7	ACC78488	Acc78488 HIV GagPo
29	2135	86.9	3564	7	ACC78489	Acc78489 HIV GagPo
30	2134.4	86.9	4716	7	ACA03522	Aca03522 Synthetic
31	2134.4	86.9	4716	9	ADC13238	Adc13238 DNA of HI
32	2130.6	86.7	3999	7	ACC78484	Acc78484 HIV GagCo
33	2129	86.7	3999	7	ACC78485	Acc78485 HIV GagCo
34	2129	86.7	3999	7	ACC78486	Acc78486 HIV GagCo
35	2129	86.7	5283	7	ACA03584	Aca03584 Synthetic
36	2129	86.7	5283	7	ACC78529	Acc78529 HIV TatRe
37	2128.8	86.6	4713	7	ACA03592	Aca03592 Synthetic
38	2128.8	86.6	4713	9	ADC13280	Adc13280 DNA of HI
39	2128.4	86.6	3462	9	ADC13237	Adc13237 DNA of HI
40	2115.2	86.1	2466	7	ACA03542	Aca03542 Synthetic
41	2115.2	86.1	2466	7	ACC78506	Acc78506 HIV p2Pol
42	2103	85.6	3735	7	ACA03545	Aca03545 Synthetic
43	2103	85.6	3735	7	ACC78509	Acc78509 HIV p2Pol
44	2096	85.3	2472	7	ACA03543	Aca03543 Synthetic
45	2096	85.3	2472	7	ACC78507	Acc78507 HIV p2Pol

ALIGNMENTS

RESULT 1
ABL39961
ID ABL39961 standard; DNA; 2457 BP.
XX
AC ABL39961;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 10; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene

CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 100.0%; Score 2457; DB 6; Length 2457;
Best Local Similarity 100.0%; Pred. No. 6.9e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db    1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db   61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db  121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db  181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db  241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db  301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db  361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db  421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db  481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db  541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db  601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db  661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db  721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db  781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db  841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db  901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db  961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
Db 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy 1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
Db 1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy 1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
Db 1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy 1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
Db 1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy 1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
Db 1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy 1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
Db 1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy 1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
Db 1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy 1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
Db 1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy 1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
Db 1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
Db 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy 1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
Db 1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980

Qy 1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
Db 1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
```

```
Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
```

RESULT 2
ACA03546
ID ACA03546 standard; DNA; 2445 BP.
XX
AC ACA03546;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 34; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 7; Length 2445;
Best Local Similarity 100.0%; Pred. No. 3.3e-294;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
```

```
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260
Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320
Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380
Qy       1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
Db       1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440
Qy       1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
Db       1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500
Qy       1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
Db       1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560
Qy       1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
Db       1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620
Qy       1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
Db       1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680
Qy       1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
Db       1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740
Qy       1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
Db       1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800
Qy       1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
Db       1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860
Qy       1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
Db       1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920
Qy       1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
Db       1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980
Qy       1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
Db       1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040
Qy       2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
Db       2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100
Qy       2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
Db       2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160
Qy       2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
Db       2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220
Qy       2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
Db       2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280
Qy       2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
Db       2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340
Qy       2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
Db       2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400
Qy       2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 3
ADC13264
ID   ADC13264 standard; DNA; 2445 BP.
XX
AC   ADC13264;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 40; 301pp; English.
```

XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 9; Length 2445;
Best Local Similarity 100.0%; Pred. No. 3.3e-294;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380

Qy       1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440

Qy       1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500

Qy       1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560

Qy       1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620

Qy       1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680

Qy       1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740

Qy       1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800

Qy       1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860

Qy       1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920
```

```
Qy       1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980

Qy       1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040

Qy       2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100

Qy       2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160

Qy       2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220

Qy       2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280

Qy       2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340

Qy       2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy       2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 4
ADC13230
ID ADC13230 standard; DNA; 3930 BP.
XX
AC ADC13230;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 6; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.4; DB 9; Length 3930;
Best Local Similarity 100.0%; Pred. No. 2.3e-293;
Matches 2437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
```

```
Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy      1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db      2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy      1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db      2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy      1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db      2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy      1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db      2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy      1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db      3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy      1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db      3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy      1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db      3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy      1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db      3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy      1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db      3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy      1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db      3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy      1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db      3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy      1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db      3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy      2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db      3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy      2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db      3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy      2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db      3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy      2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db      3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy      2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db      3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy      2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db      3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy      2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db      3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 5
ABL39960
ID   ABL39960 standard; DNA; 2463 BP.
XX
AC   ABL39960;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YM SEQ ID NO:31.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
```

PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 9; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 2.6e-293;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2455 TTC 2457
             |||
Db       2461 TTC 2463
```

```
RESULT 6
ADC13231
ID   ADC13231 standard; DNA; 3930 BP.
XX
AC   ADC13231;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagCmplPolmutAtt_C SEQ ID NO 10.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
```

```
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 7; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;
```

```
  Query Match             99.1%;  Score 2434.8;  DB 9;  Length 3930;
  Best Local Similarity   99.9%;  Pred. No. 3.6e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086
```

```
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG. 913
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746
Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806
Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866
Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926
Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986
Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046
Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106
Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166
Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226
```

```
Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286
Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346
Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406
Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466
Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526
Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586
Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646
Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706
Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766
Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826
Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886
Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 7
ADC13232
ID   ADC13232 standard; DNA; 3930 BP.
XX
AC   ADC13232;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
```

PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 8; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 99.1%; Score 2434.8; DB 9; Length 3930;
Best Local Similarity 99.9%; Pred. No. 3.6e-293;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db      1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db      1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy       134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db      1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy       194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db      1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy       254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db      1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db      1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db      1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db      1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db      1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Ov       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613

Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db      2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db      2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db      2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGGCCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy      1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db      2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy      1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db      2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy      1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db      2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy      1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db      2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy      1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db      3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy      1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
```

```
Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 8
ACA03591
ID ACA03591 standard; DNA; 5184 BP.
XX
AC ACA03591;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #74.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 79; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match 99.1%; Score 2434.8; DB 7; Length 5184;
Best Local Similarity 99.9%; Pred. No. 3.4e-293;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220
```

```
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280
Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340
Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400
Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460
Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520
Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580
Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640
Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700
Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820
Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880
Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000
Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060
Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db       4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120
Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db       4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180
Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db       4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240
Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db       4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360
Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420
Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCGCACCGCCCACACCAACGACG 1753
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCGCACCGCCCACACCAACGACG 4480
Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540
Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600
Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660
Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720
Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780
Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840
Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900
Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960
Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020
Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080
Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140
Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178

RESULT 9
ADC13279
ID   ADC13279 standard; DNA; 5184 BP.
XX
AC   ADC13279;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct TatRevNefgagCpolIna_C SEQ ID NO 58.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
```

XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ.
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 55; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match 99.1%; Score 2434.8; DB 9; Length 5184;
Best Local Similarity 99.9%; Pred. No. 3.4e-293;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db      2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db      2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy       134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db      2861 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy       194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db      2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy       254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db      2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db      3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db      3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db      3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db      3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db      3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db      3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db      3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db      3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy      1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db      4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120

Qy      1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db      4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180

Qy      1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db      4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240

Qy      1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
```

```
Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 10
ACA03547
ID ACA03547 standard; DNA; 2457 BP.
XX
AC ACA03547;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 35; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.8%; Score 2428.6; DB 7; Length 2457;
Best Local Similarity 99.6%; Pred. No. 2.2e-292;
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
```

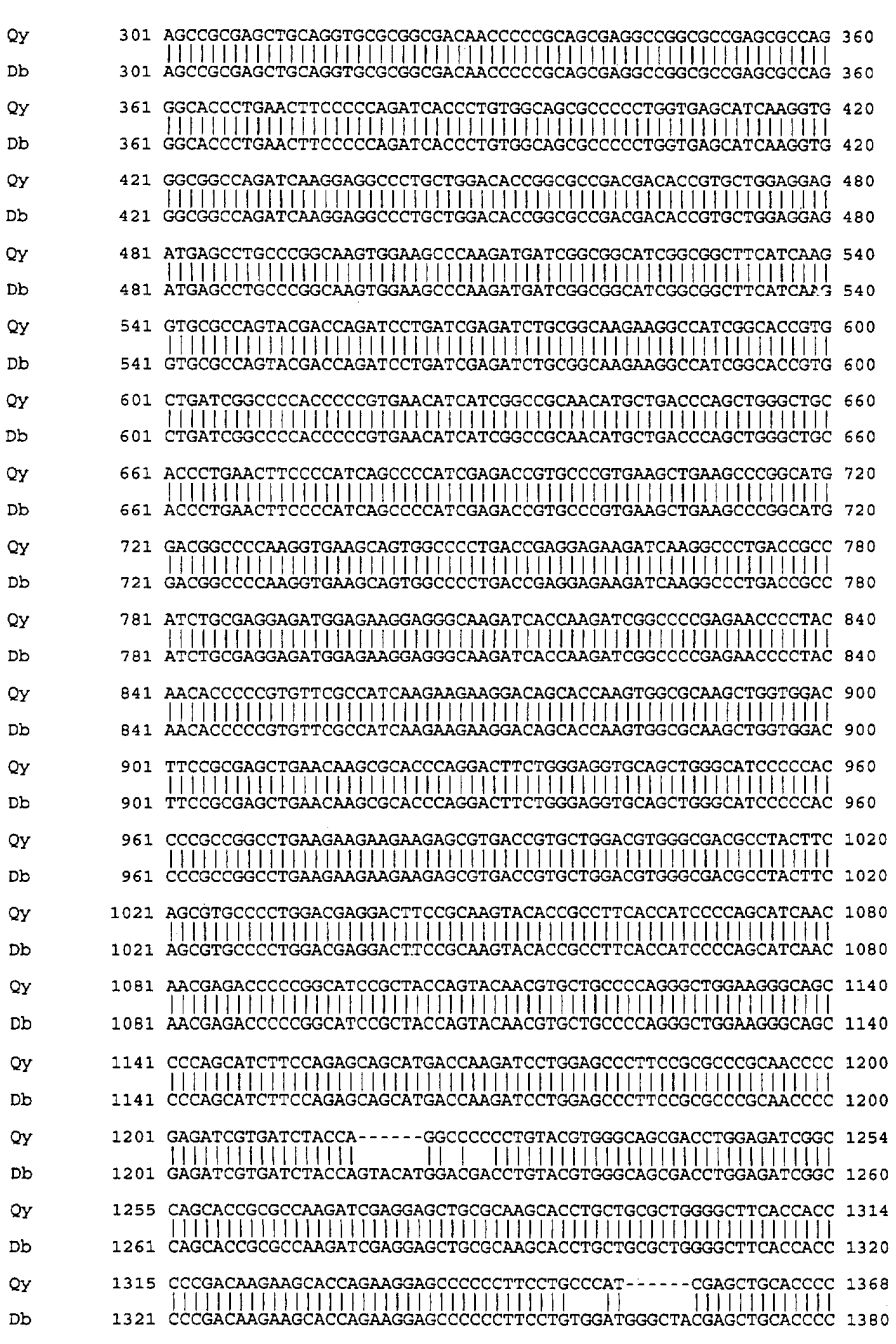
```
Qy       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
Db      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy      1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
Db      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Db      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy      1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
Db      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy      1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
Db      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy      1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
Db      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy      1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
Db      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy      1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
Db      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy      1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
Db      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy      1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
Db      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy      1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
Db      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy      1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
Db      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy      2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
Db      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy      2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
Db      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy      2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
Db      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy      2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
Db      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy      2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
Db      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy      2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
Db      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy      2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db      2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 11
ADC13265
ID ADC13265 standard; DNA; 2457 BP.
XX

```
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 41; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match             98.8%;  Score 2428.6;  DB 9;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 2.2e-292;
  Matches 2447;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              ||||||||||||||||||||||||||||||||||||    ||      ||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457

RESULT 12
ABL39959
```

```
ID   ABL39959 standard; DNA; 2469 BP.
XX
AC   ABL39959;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 8; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             98.3%;  Score 2415.4;  DB 6;  Length 2469;
  Best Local Similarity   99.3%;  Pred. No. 9.6e-291;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
```

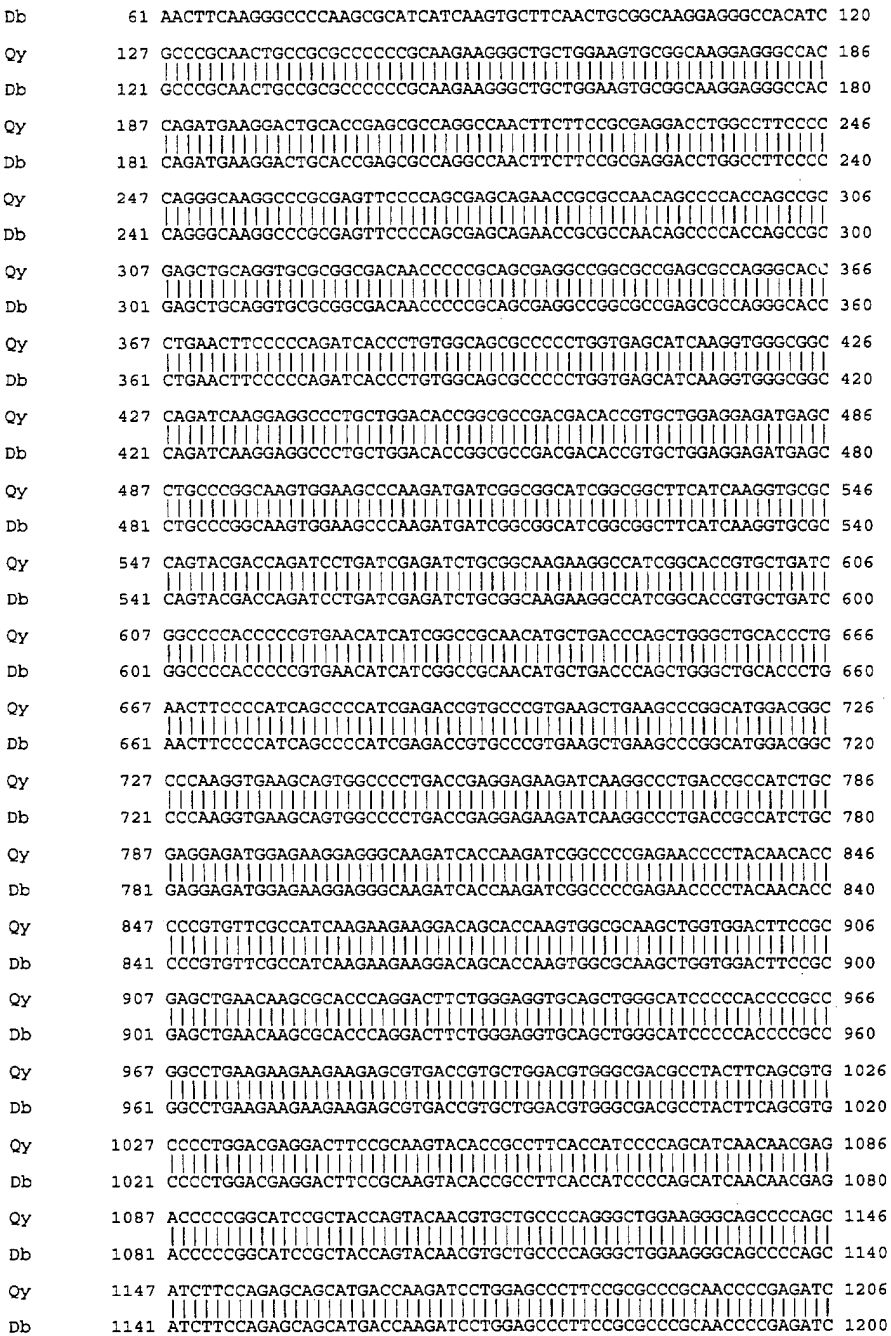

Qy 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Qy 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy 1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db 1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Qy 1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db 1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Qy 1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
Db 1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Qy 1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Qy 1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
Db 1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Qy 1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
Db 1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Qy 1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
Db 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Qy 1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
Db 1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy 1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Db 1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Qy 1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
Db 1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Qy 1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
Db 1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Qy 1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
Db 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy 1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
Db 1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Qy 1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
Db 1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Qy 2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
Db 2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Qy 2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
Db 2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Qy 2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
Db 2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Qy 2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
Db 2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Qy 2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
Db 2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Qy 2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
Db 2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Qy 2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
Db 2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Qy 2449 GGTGAATTC 2457

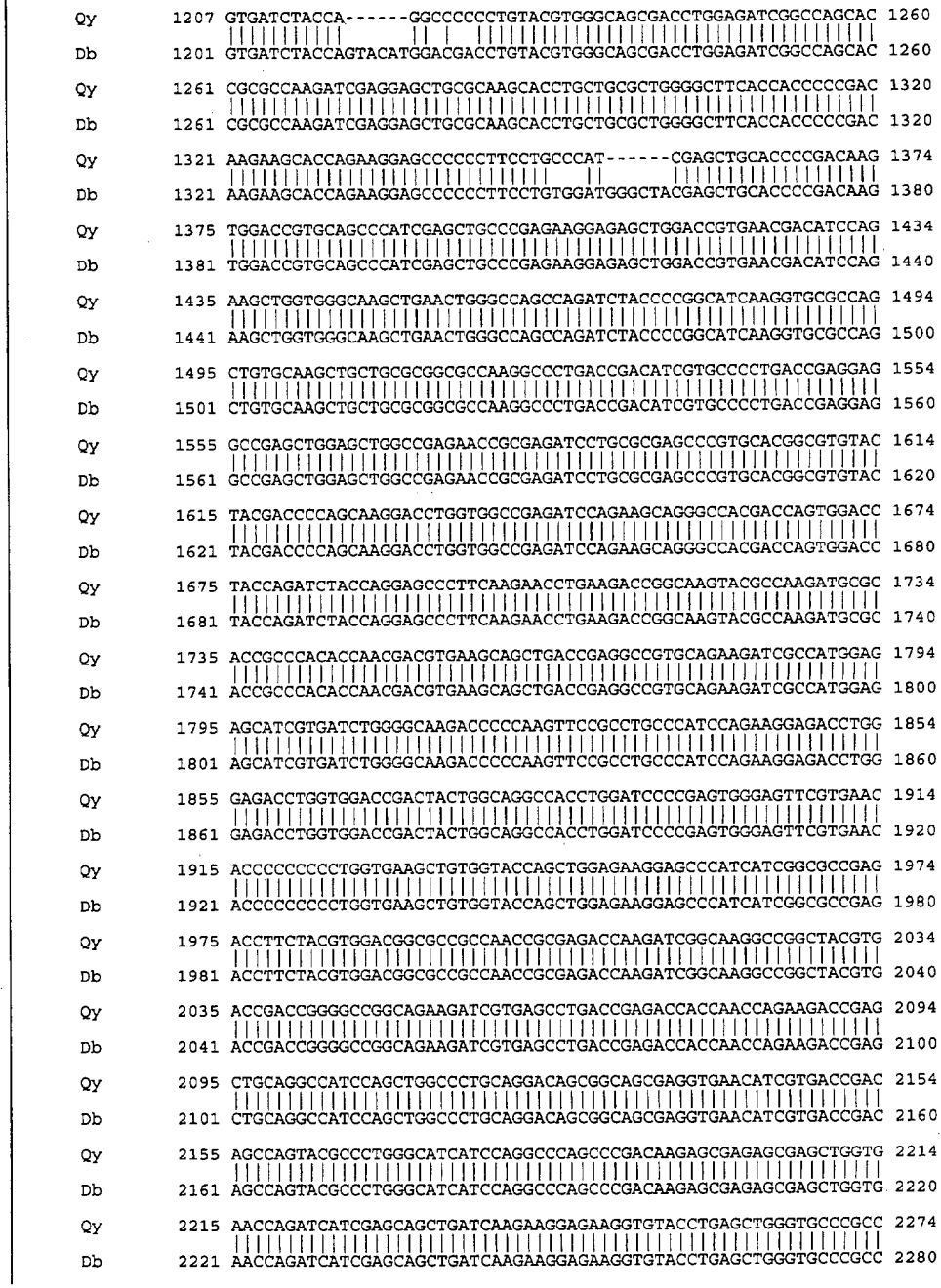
```
                 |||||||||
Db       2461 GGTGAATTC 2469


RESULT 13
ADC13234
ID   ADC13234 standard; DNA; 3531 BP.
XX
AC   ADC13234;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagPolmut_C SEQ ID NO 13.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 10; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3531 BP; 802 A; 1210 C; 1096 G; 423 T; 0 U; 0 Other;

  Query Match             97.9%;  Score 2404.4;  DB 9;  Length 3531;
  Best Local Similarity   99.1%;  Pred. No. 2.1e-289;
  Matches 2417;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| |||||||||||| || |  |   ||  ||  | || ||| |      |
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2048 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2107

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2108 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2167

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2168 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2227

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2288 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2347

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
```

```
Db       2348 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2407

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393

Db       2408 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2467

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453

Db       2468 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2527

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513

Db       2528 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2587

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573

Db       2588 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 2647

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633

Db       2648 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 2707

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693

Db       2708 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 2767

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753

Db       2768 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 2827

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813

Db       2828 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 2887

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873

Db       2888 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 2947

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933

Db       2948 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3007

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993

Db       3008 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3067

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053

Db       3068 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3127

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113

Db       3128 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3187

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173

Db       3188 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3247

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233

Db       3248 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3307

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293

Db       3308 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3367

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353

Db       3368 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3427

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413

Db       3428 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3487

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451

Db       3488 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3525
```

RESULT 14
ACA03548
ID ACA03548 standard; DNA; 2457 BP.
XX
AC ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 36; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

```
  Query Match           97.8%;  Score 2403.4;  DB 7;  Length 2457;
  Best Local Similarity  99.3%;  Pred. No. 2.9e-289;
  Matches 2439;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66

Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
```





```
Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 15
ADC13266
ID ADC13266 standard; DNA; 2457 BP.
XX
AC ADC13266;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 42; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 97.8%; Score 2403.4; DB 9; Length 2457;
Best Local Similarity 99.3%; Pred. No. 2.9e-289;
Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
             |||||||||||      || |  ||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
             |||||||||||||||||||||||||||||||    ||      ||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

Search completed: April 10, 2004, 07:32:35
Job time : 626.487 secs